

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.

« 29 » августа 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1. В.16 Обработка данных на Python

Направление подготовки 01.03.02 Прикладная математика и информатика

Профиль Современные методы машинного обучения и компьютерного зрения

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины «Обработка данных на Python» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 01.03.02 Прикладная математика и информатика

Программу составил(и):

Харченко Анна Владимировна, доцент, канд. пед наук

И.О. фамилия, должность, ученая степень, ученое звание



подпись

Рабочая программа дисциплины утверждена на заседании центра
искусственного интеллекта
протокол № 01 «28» августа 2025 г.
Руководитель центра ИИ Коваленко А.В.



подпись

Утверждена на заседании учебно-методической комиссии факультета
компьютерных технологий и прикладной математики
протокол № 01 «28» августа 2025 г.
Председатель УМК факультета Коваленко А.В.



подпись

Рецензенты:

Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

1 Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины

Цель дисциплины — сформировать у студентов базовые компетенции в области обработки и анализа данных с использованием языка программирования Python. В ходе освоения дисциплины студенты изучат фундаментальные принципы работы с различными типами данных и форматами, освоят основные библиотеки для их загрузки, очистки, преобразования и первичного визуального анализа. Это позволит создать прочную основу для дальнейшего изучения более сложных дисциплин, таких как машинное обучение и глубокое обучение, и сформировать навыки, необходимые для решения практических задач анализа данных.

1.2 Задачи дисциплины

В рамках изучения дисциплины решаются следующие задачи:

- Изучение основ языка Python и его экосистемы библиотек для анализа данных (NumPy, Pandas, Matplotlib, Seaborn).
- Освоение методов загрузки данных из различных источников и форматов (CSV, JSON, таблицы).
- Приобретение практических навыков проверки данных на корректность, выявления и обработки пропусков, дубликатов и аномалий.
- Формирование умения выполнять первичный анализ и визуализацию данных для выявления закономерностей и тенденций.
- Изучение базовых методов предварительной обработки данных, включая преобразование и кодирование признаков.
- Освоение принципов работы с текстовыми данными: их очистка, токенизация и векторизация.
- Формирование понимания принципов формирования и подготовки качественных наборов данных (datasets).

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Обработка данных на Python» относится к «Вариативная часть» Блока 1 «Дисциплины (модули)» учебного плана.

1.4 Профессиональные роли в структуре образовательной программы

Роль 1: Data Analyst (Аналитик данных)

Извлечение знаний из данных, построение аналитических моделей, использующих МО и ИИ.

Задачи:

- 1. Статистический анализ, визуализация данных, предварительная обработка.*
- 2. Создание прогнозных моделей*
- 3. Построение аналитических моделей для поддержки бизнес-решений.*

Роль 2: MLOps (Специалист по эксплуатации ИИ)

Автоматизация и операционное управление жизненным циклом МО-моделей

Задачи:

- 1. DevOps для ML.*
- 2. Автоматизация, мониторинг ML-систем.*
- 3. Операционное управление жизненным циклом ML-моделей.*

Роль 3: AI PM (Менеджер проектов ИИ)

Управление процессами создания ИИ-решений, включая координацию команды разработки

Задачи:

- 1. Управление ИИ-проектами от идеи до внедрения*

2. Анализ бизнес-требований и постановка задач

3. Оценка эффективности и ROI ИИ-решений

1.5 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

BD-1	Способен осуществлять поиск, сбор, очистку и предварительный анализ данных
BD-1.4	Отбирает признаки данных, значимые для исследования
Знать	основные типы данных, критерии оценки информативности признаков, методы и алгоритмы отбора признаков.
Уметь	анализировать распределения признаков, применять методы <i>feature selection</i> , интерпретировать результаты.
Владеть	навыками использования специализированного программного обеспечения и библиотек для анализа и отбора признаков.
BD-2	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных
BD-2.2	Работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность
Знать	форматы хранения данных, принципы работы с базами данных и сетевыми API, типовые проблемы качества данных.
Уметь	загружать данные из различных источников, проводить первичный осмотр выявлять аномалии, осуществлять базовую очистку данных.
Владеть	технологиями извлечения, преобразования и загрузки данных на базовом уровне.
ML-2	Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками
ML-2.2	Применяет методы предварительной обработки данных и работы с признаками
Знать	методы обработки пропусков, кодирования категориальных масштабирования числовых признаков, концепцию конвейеров обработки данных.
Уметь	выбирать и применять методы предобработки в зависимости от типа данных, строить конвейеры обработки
Владеть	навыками применения библиотек для предварительной обработки данных и создания пайплайнов.
LLM-2	Дообучение и адаптация генеративных моделей
LLM-2.2	Создаёт обучающие наборы данных

Знать		<i>принципы разметки данных, методы аугментации, требования к качеству datasets.</i>
Уметь		<i>формировать репрезентативные выборки, применять аугментацию, организовывать процесс разметки.</i>
Владеть		<i>навыками использования инструментов для аугментации, методами проверки согласованности разметки, технологиями версионирования данных.</i>

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зач. ед. (72 часф), их распределение по видам работ представлено в таблице

Вид учебной работы	Всего часов	Семестры (часы)					
		3					
Контактная работа, в том числе:	52,2	52,2					
Аудиторные занятия (всего):							
Занятия лекционного типа	16	16					
Лабораторные занятия	34	34					
Иная контактная работа:	2,2	2,2					
Контроль самостоятельной работы (КСР)	2	2					
Промежуточная аттестация (ИКР)	0,2	0,2					
Самостоятельная работа, в том числе:	19,8	19,8					
Проработка учебного (теоретического) материала	2	2					
<i>Выполнение индивидуальных заданий (подготовка сообщений, презентаций)</i>	15	15					
Подготовка к текущему контролю	4,8	4,8					
Общая трудоёмкость	час.	72	72				
	в том числе контактная работа	52,2	52,2				
	зач. ед	2	2				

2.2 Структура дисциплины

Распределение видов учебной работы и их трудоёмкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 3 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Основы работы с данными в python	14	4		6	4
2.	Базовые операции и анализ данных	16	4		8	4

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
3.	Визуализация и исследовательский анализ	12	2		6	4
4.	Преобразование и обогащение данных	14	2		8	4
5.	Интеграция и автоматизация обработки данных	13,8	4		6	3,8
ИТОГО по разделам дисциплины		69,8	16		34	19,8
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,2				
Подготовка к текущему контролю						
Общая трудоемкость по дисциплине		72				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Основы работы с данными в python	Введение в экосистему Python для обработки данных. Роль языка Python в анализе данных. Обзор ключевых библиотек: NumPy для численных вычислений, Pandas для работы со структурированными данными, Matplotlib и Seaborn для визуализации. Понятие DataFrame и Series. Установка и настройка рабочего окружения (Jupyter Notebook)	К, РЗ
2.	Основы работы с данными в python	Начало работы с данными: загрузка и первичный осмотр. Методы загрузки данных из CSV и Excel-файлов с использованием библиотеки Pandas. Базовые методы первичного исследования данных: head(), info(), describe(). Понимание структуры данных, типов переменных и поиск пропущенных значений.	К, РЗ

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
3.	Базовые операции и анализ данных	Базовые операции с DataFrame в Pandas. Индексация и выбор данных (loc, iloc). Фильтрация строк по условию. Создание, модификация и удаление столбцов. Сортировка данных. Агрегация данных с помощью методов groupby() и agg().	К, РЗ
4.	Базовые операции и анализ данных	Предварительная обработка данных: работа с пропусками и дубликатами. Стратегии обработки пропущенных данных: обнаружение (isnull()), удаление (dropna()), заполнение (fillna()). Методы выявления и удаления дубликатов. Принципы обеспечения целостности и чистоты данных на практических примерах.	К, РЗ
5.	Визуализация и исследовательский анализ	Визуализация данных для анализа и исследования. Построение базовых графиков с использованием Matplotlib и Seaborn: гистограммы, столбчатые диаграммы, диаграммы рассеяния. Инструменты для визуального исследования распределений и взаимосвязей между переменными.	К, РЗ
6.	Преобразование и обогащение данных	Основы работы с текстовыми данными в Pandas. Работа с текстовыми столбцами в DataFrame. Базовые строковые методы в Pandas (.str.lower(), .str.replace(), .str.split()). Поиск и фильтрация данных по текстовым шаблонам. Примеры очистки и предварительной обработки текстовой информации.	К, РЗ
7.	Интеграция и автоматизация обработки данных	Сборка пайплайна обработки данных и итоговое повторение. Обобщение изученного материала. Построение сквозного примера (пайплайна) обработки данных: от загрузки сырых данных до получения чистого набора, готового для анализа. Обсуждение лучших практик и типичных ошибок.	К, РЗ
8.	Интеграция и автоматизация обработки данных	Введение в автоматизацию и создание отчетов.	К, РЗ

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		Основы автоматизации процессов обработки данных: создание скриптов на Python. Генерация сводных таблиц и отчетов средствами Pandas. Экспорт результатов анализа в файлы. Краткий обзор направлений для дальнейшего изучения.	

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.2 Занятия семинарского типа

Не предусмотрены

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.3 Лабораторные занятия

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Основы работы с данными в python	Знакомство с рабочим окружением. Установка и настройка Python, Jupyter Notebook. Основы синтаксиса Python: переменные, типы данных, структуры (списки, словари). Работа с библиотеками (import).	ЛР
2.	Основы работы с данными в python	Начало работы с Pandas: загрузка и первичный осмотр данных. Загрузка данных из CSV-файла в DataFrame. Использование методов head(), tail(), info(), describe() для первичного анализа. Навигация по DataFrame.	ЛР
3.	Основы работы с данными в python	Базовые операции с данными в Pandas. Индексация и выбор данных с помощью loc и iloc. Создание новых столбцов. Сортировка данных по значениям столбца.	ЛР
4.	Базовые операции и анализ данных	Фильтрация и агрегация данных. Фильтрация строк по одному и нескольким условиям. Группировка данных с помощью groupby().	ЛР

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
		Применение агрегирующих функций (sum(), mean(), count()).	
5.	Базовые операции и анализ данных	Объединение данных из нескольких источников. Конкатенация DataFrame. Соединение таблиц по индексу и столбцам с использованием merge().	ЛР
6.	Базовые операции и анализ данных	Поиск и обработка пропущенных значений. Поиск пропусков с помощью isnull(). Стратегии обработки: удаление строк/столбцов (dropna()), заполнение константой или медианой (fillna()).	ЛР
7.	Базовые операции и анализ данных	Работа с дубликатами и аномалиями. Поиск и удаление полных дубликатов (duplicated(), drop_duplicates()). Поиск аномалий с использованием описательных статистик и визуализации (boxplot).	ЛР
8.	Визуализация и исследовательский анализ	Визуализация данных: основы Matplotlib. Построение линейных графиков, гистограмм и диаграмм рассеяния. Настройка подписей осей и заголовков.	ЛР
9.	Визуализация и исследовательский анализ	Визуализация данных: библиотека Seaborn. Построение статистических графиков: barplot, histplot, scatterplot. Создание визуализаций для анализа распределений и взаимосвязей.	ЛР
10.	Визуализация и исследовательский анализ	Создание сводных таблиц. Содержание: Построение сводных таблиц с помощью метода pivot_table() для агрегации и анализа данных.	ЛР
11.	Преобразование и обогащение данных	Преобразование данных: работа с датами и временем. Содержание: Преобразование строк в формат datetime. Извлечение компонентов даты (год, месяц, день). Расчет разниц между датами.	ЛР
12.	Преобразование и обогащение данных	Преобразование данных: кодирование категориальных признаков. Применение методов Label Encoding и One-Hot Encoding для преобразования текстовых категориальных данных в числовой формат.	ЛР
13.	Преобразование и обогащение данных	Работа с текстовыми данными в столбцах DataFrame.	ЛР

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
		Применение строковых методов Pandas (.str): приведение к регистру, замена подстрок, разделение текста на части. Поиск подстрок.	
14.	Преобразование и обогащение данных	Работа с категориальными данными Закодируйте категориальные переменные разными методами (Label Encoding, One-Hot Encoding), сравните результаты преобразования.	ЛР
15.	Интеграция и автоматизация обработки данных	Автоматизация обработки: написание функций и их применение к DataFrame. Создание пользовательских функций. Применение функций к столбцам и строкам DataFrame с использованием apply() и applymap().	ЛР
16.	Интеграция и автоматизация обработки данных	Сквозной проект (часть 1). Решение комплексной задачи: загрузка сырых данных, их очистка, обработка пропусков и дубликатов, первичный анализ	ЛР
17.	Интеграция и автоматизация обработки данных	Сквозной проект (часть 2). Продолжение проекта: преобразование данных, углубленный анализ, визуализация ключевых инсайтов и формирование итогового отчета.	ЛР

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.4 Примерная тематика курсовых работ (проектов)

Не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Изучение теоретического материала	Методические указания по организации самостоятельной работы студентов, утвержденные кафедрой информационных технологий, протокол №1 от 30.08.2019
2	Решение задач	Методические указания по организации самостоятельной работы студентов, утвержденные кафедрой информационных технологий, протокол №1 от 30.08.2019

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

– Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

– Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

– Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

– Информационно-коммуникационные технологии (ИКТ) - расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

– Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

– Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

– Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

– Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

– Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

- Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

- Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

- работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;

- проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;

- анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;

- развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Семестр	Вид занятия	Используемые интерактивные образовательные технологии	количество интерактивных часов
3	ЛР	Практические занятия в режимах взаимодействия «преподаватель – студент» и «студент – студент»	10
Итого			10

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4. Оценочные и методические материалы

4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Методы программирования».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме разноуровневых заданий, и **промежуточной аттестации** в форме вопросов и заданий к зачету и экзамену.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,

– в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Контролируемые разделы (темы) дисциплины*	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	Основы работы с данными в python	BD-1.4; BD-2.2; ML-2.2; LLM-2.2	Вопросы к зачету: 1-4	Типовые задачи: 1

2	Базовые операции и анализ данных	BD-1.4; BD-2.2; ML-2.2; LLM-2.2	Вопросы к зачету: 5,6,9-11, 18	Типовые задачи: 2-4,8,12,17
3	Визуализация и исследовательский анализ	BD-1.4; BD-2.2; ML-2.2; LLM-2.2	Вопросы к зачету: 12,13	Типовые задачи: 6,7
4	Преобразование и обогащение данных	BD-1.4; BD-2.2; ML-2.2; LLM-2.2	Вопросы к зачету: 15	Типовые задачи: 9
5	Интеграция и автоматизация обработки данных	BD-1.4; BD-2.2; ML-2.2; LLM-2.2	Вопросы к зачету 7,8, 14,16,17	Типовые задачи 5, 10, 11, 13-16

Показатели, критерии и шкала оценки сформированных компетенций

Соответствие **пороговому уровню** освоения компетенций планируемым результатам обучения и критериям их оценивания (оценка: **удовлетворительно /зачтено**):

BD-1.4	Отбирает признаки данных, значимые для исследования	
Знать		<i>основные типы данных (числовые, категориальные, текстовые) и методы их первичного анализа.</i>
Уметь		<i>выполнять первичный анализ признаков (расчет базовых статистик, построение графиков распределения) с помощью библиотек Pandas и Matplotlib/Seaborn; отбирать признаки на основе анализа корреляций.</i>
Владеть		<i>навыками использования методов .describe(), .value_counts(), .corr() в Pandas; основами визуализации данных для анализа признаков.</i>
BD-2	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных	
BD-2.2	Работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность	
Знать		<i>основные структуры данных в Python (списки, словари) и структуры данных библиотеки Pandas (Series, DataFrame); основные форматы файлов (CSV, JSON, XLSX) и способы их чтения.</i>
Уметь		<i>загружать данные из файлов различных форматов (CSV, JSON) в объекты Pandas DataFrame; проверять данные на наличие пропусков, дубликатов и аномальных значений; выполнять базовые операции по очистке данных (удаление дубликатов, заполнение пропусков базовыми методами).</i>
Владеть		<i>технологиями чтения данных с помощью pandas.read_csv(), pandas.read_json(); методами первичной проверки данных (.info(), .isnull().sum(), .duplicated().sum()).</i>

ML-2	Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками	
<i>ML-2.2</i>		<i>Применяет методы предварительной обработки данных и работы с признаками</i>
Знать		<i>базовые методы предобработки данных: масштабирование числовых признаков, кодирование категориальных признаков (One-Hot Encoding, Label Encoding).</i>
Уметь		<i>применять классы StandardScaler, MinMaxScaler, OneHotEncoder из библиотеки Scikit-learn для преобразования признаков.</i>
Владеть		<i>навыками использования трансформеров Scikit-learn для подготовки данных к дальнейшему анализу.</i>
LLM-2	Дообучение и адаптация генеративных моделей	
<i>LLM-2.2</i>		<i>Создаёт обучающие наборы данных</i>
Знать		<i>основные этапы и принципы создания датасетов для анализа данных; требования к качеству данных (полнота, непротиворечивость).</i>
Уметь		<i>выполнять базовую разметку и категоризацию данных; формировать выборки данных из различных источников.</i>
Владеть		<i>навыками формирования и сохранения datasets в форматах CSV/JSON.</i>

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Типовые задачи для самостоятельного решения:

1. Загрузите данные из CSV-файла, определите размер DataFrame, выведите первые 7 строк, получите основную статистику для числовых столбцов.
2. Для заданного набора данных определите столбцы с пропущенными значениями, визуализируйте количество пропусков, заполните пропуски в числовых столбцах медианными значениями.
3. Отфильтруйте DataFrame по условию (например, товары дороже 1000 руб.), отсортируйте результат по убыванию цены, сохраните в новый файл.
4. Найдите полные дубликаты в данных, удалите их, оставив первое вхождение. Проверьте наличие частичных дубликатов по ключевым полям.
5. Сгруппируйте данные по категориям, для каждой категории посчитайте среднее значение, количество элементов и стандартное отклонение целевого показателя.
6. Постройте гистограмму распределения возраста клиентов и коробчатую диаграмму для анализа выбросов в данных о доходах.
7. Создайте матрицу корреляций для числовых признаков, визуализируйте ее с помощью тепловой карты. Постройте диаграммы рассеяния для наиболее коррелирующих пар.

8. На основе существующих столбцов создайте 2-3 новых признака (например, вычисление возраста из даты рождения, категоризация по ценовым сегментам).
9. Для столбца с текстовыми отзывами выполните предобработку: приведите к нижнему регистру, удалите пунктуацию и стоп-слова.
10. Преобразуйте строковый столбец с датами в формат datetime, выделите месяц и год, постройте график динамики показателя по месяцам.
11. Создайте сводную таблицу, показывающую среднюю выручку по регионам и товарным категориям.
12. Найдите выбросы в данных с помощью метода IQR, замените их граничными значениями или удалите соответствующие строки.
13. Объедините два DataFrame по общему ключу, выполните разные типы соединений (inner, left), проанализируйте различия в результатах.
14. Напишите функцию для очистки и предобработки DataFrame, применяющую стандартные операции (удаление дубликатов, заполнение пропусков, etc.).
15. Проведите полный анализ набора данных: от первичной загрузки до визуализации ключевых закономерностей, оформив результаты в виде структурированного отчета.
16. Закодируйте категориальные переменные разными методами (Label Encoding, One-Hot Encoding), сравните результаты преобразования.
17. Проанализируйте производительность операций с DataFrame, примените методы оптимизации (изменение типов данных, использование векторных операций).
18. Напишите проверки на корректность данных: соответствие типов, допустимые диапазоны значений, проверка уникальности ключевых идентификаторов.

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен/зачет)

Вопросы для подготовки к зачету

1. Роль языка Python и его основных библиотек (Pandas, NumPy, Matplotlib) в обработке и анализе данных.
2. Основные структуры данных библиотеки Pandas (Series и DataFrame). Их свойства и отличия.
3. Методы загрузки данных из внешних файлов (CSV, Excel) в Pandas DataFrame. Ключевые параметры функций чтения.
4. Методы первичного исследования данных в Pandas: head(), info(), describe(). Какую информацию позволяет получить каждый из них?
5. Индексация и выбор данных в DataFrame. Различия между методами loc и iloc.
6. Методы фильтрации строк в DataFrame по одному или нескольким условиям.
7. Назначение и использование метода groupby() в Pandas. Примеры применения агрегирующих функций.
8. Процесс объединения данных из нескольких DataFrame с помощью метода merge(). Типы соединений (inner, outer, left, right).
9. Поиск и обработка пропущенных значений (NaN) в DataFrame. Методы isnull(), dropna(), fillna().
10. Методы поиска и удаления дубликатов в DataFrame (duplicated(), drop_duplicates()).

11. Создание и модификация столбцов в DataFrame. Примеры вычисляемых столбцов.
12. Базовые методы визуализации данных: гистограмма, диаграмма рассеяния, столбчатая диаграмма. В каких случаях применяется каждый тип графика?
13. Различия между библиотеками визуализации Matplotlib и Seaborn. Преимущества Seaborn для статистической графики.
14. Методы преобразования категориальных данных: Label Encoding и One-Hot Encoding. Их различия и сферы применения.
15. Работа с текстовыми данными в столбцах DataFrame. Основные строковые методы из набора .str в Pandas.
16. Создание сводных таблиц с помощью метода pivot_table(). Основные параметры метода.
17. Назначение и применение метода apply() для DataFrame и Series. Пример создания пользовательской функции.
18. Основные этапы подготовки данных (Data Preprocessing) перед анализом.

4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Методические рекомендации, определяющие процедуры оценивания выполнения контрольных заданий:

Задание считается выполненным при выполнении следующих условий:

- предоставлен исходный код на C++ в среде Visual Studio
- продемонстрирована работоспособность программы
- студент понимает исходный код и отвечает на вопросы по его организации.

Методические рекомендации к сдаче зачета

Процедура промежуточной аттестации проходит в соответствии с Положением о текущем контроле и промежуточной аттестации обучающихся ФГБОУ ВО «КубГУ».

Текущий контроль успеваемости студентов проводится в целях совершенствования и непрерывного контроля качества образовательного процесса, проверки усвоения учебного материала, активизации самостоятельной работы студентов, стимулирования их учебной работы, обеспечения эффективности образовательного процесса, предупреждения рисков отчисления студентов.

Текущий контроль знаний студентов осуществляется постоянно в течение всего семестра.

Виды текущего контроля: устный (письменный) опрос на занятиях; проверка выполнения домашних заданий; проведение контрольных работ; оценка активности студента на занятии.

Студенты очной формы обучения обязаны сдать зачет до начала экзаменационной сессии.

По результатам сдачи зачета выставляется оценка «зачтено»/ «не зачтено».

Выставление зачетов для студентов очной формы обучения проводятся в период до экзаменационной сессии.

При отсутствии зачетной книжки у студента экзаменатор не имеет права принимать у него зачет/экзамен. Такой студент считается не явившимся на зачет/экзамен. В

исключительных случаях, на основании распоряжения декана преподаватель может допустить студента к зачету/экзамену при наличии документа, удостоверяющего личность.

В целях объективного оценивания знаний во время проведения зачетов и экзаменов не допускается наличие у студентов посторонних предметов и технических устройств.

Студенту, использующему в ходе зачета неразрешенные источники и средства получения информации, выставляется неудовлетворительная оценка, и он удаляется из аудитории.

Во время зачета студенты могут пользоваться утвержденной рабочей программой учебной дисциплины, которая должна быть в наличии на экзамене, а также с разрешения экзаменатора справочной литературой и другими пособиями.

Студенты, нарушающие правила поведения при проведении зачетов и экзаменов, могут быть незамедлительно удалены из аудитории, к ним могут быть применены меры дисциплинарного воздействия.

На зачете/экзамене могут присутствовать ректор, проректор по учебной работе, декан факультета, заведующий кафедрой, которая обеспечивает учебный процесс по данной дисциплине. Присутствие на экзаменах и зачетах посторонних лиц без разрешения ректора или проректора по учебной работе не допускается.

После прослушивания лекции рекомендуется выполнить упражнения, приводимые в лекции для самостоятельной работы, а также выполнить на компьютере с использованием среды Python задачи, приводимые в лекции в качестве примеров.

При самостоятельной работе студентов необходимо изучить литературу, приведенную в перечнях, для осмысления вводимых понятий, анализа предложенных подходов и методов разработки параллельных программ. Разрабатывая решение новой задачи, студент должен уметь выбрать методы решения задачи с учетом целевой аппаратной платформы, проводить отладку и профилирование программы.

Критерии оценивания

Оценка	
Не зачтено	Зачтено
<ul style="list-style-type: none">• если студент правильно решил менее 70 % задач и/или не имеет представление как решать остальные задачи• не знает значительной части теоретического материала, допускает существенные ошибки.	<ul style="list-style-type: none">• если студент правильно решил 70 % задач, имеет представление как решать остальные задачи• на теоретические вопросы дан развернутый ответ. Материал изложен в целом последовательно. Имеются логичные и аргументированные выводы

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,

– в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4.3. Методические указания по организации вычислительной инфраструктуры

Условия применения:

- Курс рассчитан на студентов 2 курса.
- Наличие персональных компьютеров с установленным Python 3.9+ с необходимыми библиотеками.
- Доступ к среде разработки Jupyter Notebook/Lab или PyCharm/VSCode.
- Использование облачных решений (Google Colab, Yandex Cloud) как дополнительной опции.

Цели, задачи и ожидаемые результаты:

Цели:

- Обеспечить студентов стабильной и унифицированной средой для выполнения практических заданий.
- Сформировать навыки работы с основными инструментами обработки данных в Python.

Задачи преподавателя:

- Разработка руководства по настройке рабочего окружения.
- Подготовка шаблонов Jupyter Notebook для лабораторных работ.
- Создание тестовых наборов данных для отработки практических навыков.

Ожидаемые результаты студентов:

- Умение самостоятельно настраивать рабочее окружение для анализа данных.
- Навыки работы с Jupyter Notebook и основными библиотеками Python для анализа данных.
- Понимание принципов организации вычислительного процесса при работе с данными.

Порядок реализации:

Настройка окружения:

- Для локальной установки: использование Anaconda Distribution с предустановленными библиотеками (pandas, numpy, matplotlib, seaborn, scikit-learn).
- Для облачной работы: настройка доступа к Google Colab с mounting Google Drive для хранения данных.

Шаблоны работ:

Стандартная структура Jupyter Notebook для лабораторных работ:

1. Постановка задачи
2. Импорт библиотек

3. Загрузка данных
4. Предобработка и анализ
5. Визуализация результатов
6. Выводы

Порядок проверки корректности:

Чек-лист работоспособности окружения:

- Успешный запуск Jupyter Notebook
- Корректный импорт всех необходимых библиотек
- Возможность загрузки тестовых datasets
- Отсутствие ошибок при выполнении базовых операций с данными

4.4. Методические указания по организации лабораторных работ

Условия применения:

- Наличие настроенного рабочего окружения у всех студентов.
- Доступ к методическим материалам и наборам данных.
- Наличие системы контроля версий (Git) для отслеживания прогресса.

Цели, задачи и ожидаемые результаты:

Цели:

- Закрепить теоретические знания на практике.
- Сформировать устойчивые навыки обработки и анализа данных.

Задачи преподавателя:

- Разработка пошаговых руководств к лабораторным работам.
- Подготовка наборов данных различной сложности.
- Организация системы проверки и обратной связи.

Ожидаемые результаты студентов:

- Умение применять библиотеки Python для решения практических задач анализа данных.
- Навыки интерпретации результатов анализа.
- Способность оформлять результаты в соответствии с требованиями.

Порядок реализации:

План лабораторных работ:

ЛР1: Настройка окружения и первичный анализ данных

ЛР2: Базовые операции с DataFrame в Pandas

ЛР3: Визуализация данных средствами Matplotlib и Seaborn

ЛР4: Обработка пропущенных значений и дубликатов

ЛР5: Группировка и агрегация данных

ЛР6: Работа с текстовыми данными в Pandas

ЛР7: Создание сводных таблиц и отчетов

ЛР8: Сквозной проект анализа данных

Пример индивидуального задания (ЛР4):

Задача: Провести полную предобработку набора данных с пропусками и дубликатами

Критерии оценки: Полнота обработки, корректность применяемых методов, качество выводов

Контрольные вопросы:

- Методы обнаружения пропущенных значений
- Стратегии обработки пропусков различных типов
- Алгоритм поиска и удаления дубликатов

Критерии оценки:

Зачтено: Все этапы обработки выполнены корректно, сформированы обоснованные выводы, студент может объяснить примененные методы.

Не зачтено: Допущены существенные ошибки в обработке данных, студент не понимает принципы применяемых методов.

4.5. Методические указания по организации проектной деятельности студентов

Условия применения:

- Курс рассчитан на студентов 2 курса.
- Время на проект - 10-15 часов на команду.
- Наличие базовых навыков работы с Python и библиотеками анализа данных.

Цели, задачи и ожидаемые результаты:

Цели:

- Применить полученные знания для решения комплексной задачи анализа данных.
- Развить навыки командной работы и проектного подхода.

Задачи преподавателя:

- Формирование тематики проектов, соответствующих уровню студентов.
- Разработка четких технических заданий.
- Организация промежуточных контролей и финальной защиты проектов.

Ожидаемые результаты студентов:

- Сформированный навык работы с реальными данными.
- Умение презентовать результаты анализа.
- Опыт работы в команде над Data Science проектом.

Порядок реализации:

Примеры проектов:

- Анализ рынка недвижимости города
- Исследование потребительской корзины
- Анализ успеваемости студентов
- Исследование рынка труда IT-специалистов

ТЗ для проекта "Анализ рынка недвижимости":

- Загрузить данные из CSV-файла
- Провести полную предобработку данных
- Выполнить разведочный анализ
- Сформировать 5 ключевых инсайтов о рынке
- Подготовить презентацию с результатами

Критерии оценки:

Зачтено: Проект реализован в соответствии с ТЗ, продемонстрировано понимание методов анализа, сформированы содержательные выводы.

Не зачтено: Проект не соответствует базовым требованиям ТЗ, отсутствует понимание примененных методов.

Порядок проверки корректности:

- Соответствие проекта техническому заданию
- Качество кода и визуализаций
- Глубина анализа и обоснованность выводов
- Качество презентации результатов

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

5.1 Основная литература:

1. Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2024. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/544190> (дата обращения: 08.06.2025).

2. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для среднего профессионального образования / Д. Ю. Федоров. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2021. — 210 с. — (Профессиональное образование). — ISBN 978-5-534-12829-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/487079> (дата обращения: 08.06.2025).

3. Гниденко, И. Г. Технология разработки программного обеспечения : учебное пособие для среднего профессионального образования / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2024. — 248 с. — (Профессиональное образование). — ISBN 978-5-534-18131-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/539215> (дата обращения: 08.06.2025).

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Библиоклуб» и «Юрайт».

5.2 Дополнительная литература:

1. Сузи, Р. А. Язык программирования Python : учебное пособие : [16+] / Р. А. Сузи. — 2-е изд., испр. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ) : Бином. Лаборатория знаний, 2007. — 327 с. — (Основы информационных технологий). — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=233288> (дата обращения: 08.06.2025). — ISBN 978-5-9556-0109-0. — Текст : электронный.

2. Шелудько, В. М. Язык программирования высокого уровня Python: функции, структуры данных, дополнительные модули : [16+] / В. М. Шелудько. — Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. — 108 с. : ил. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=500060> (дата обращения: 08.06.2025). — Библиогр. в кн. — ISBN 978-5-9275-2648-2. — Текст : электронный.

3. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python : [16+] / В. М. Волкова, М. А. Семенова,

Е. С. Четвертакова, С. С. Вожов. – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576496> (дата обращения: 08.06.2025). – Библиогр.: с. 48. – ISBN 978-5-7782-3183-2. – Текст : электронный.

5.3. Периодические издания:

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
2. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

5.4. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
7. Президентская библиотека им. Б.Н. Ельцина <https://www.prlib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;

4. Федеральный портал "Российское образование" <http://www.edu.ru/>;
5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.uceba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--plai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru>;
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля)

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Кейсы ПАО «Сбербанк»

Кейс 1: Предобработка данных для анализа инвестиционных обзоров

Описание: Необходимо подготовить к анализу структурированные данные: биржевые котировки, макроэкономические показатели, рыночные события. Данные содержат пропуски, разноформатные даты, требуют нормализации.

Цель: Выполнить полный цикл предобработки данных для последующего анализа.

Ожидаемый результат: Очищенный dataset в формате, пригодном для анализа. Отчет о проведенных преобразованиях данных.

Кейс 2: Анализ структурированных данных клиентских обращений

Описание: Имеется dataset с обращениями клиентов, содержащий структурированные поля: категория проблемы, временные метки, статус обработки. Требуется выявить частотность различных типов проблем, сезонность обращений.

Цель: Провести разведочный анализ данных (EDA) для выявления основных "болевых точек".

Ожидаемый результат: Дашборд с визуализацией основных метрик и статистик по обращениям.

Кейс 3: Подготовка синтетических банковских данных

Описание: Требуется создать синтетический dataset, имитирующий банковские транзакции, с сохранением реальных статистических закономерностей (распределения сумм, временные паттерны).

Цель: Освоить методы генерации синтетических данных и их верификации.

Ожидаемый результат: Синтетический dataset и отчет о сравнении его статистических характеристик с реальными данными.

Кейсы от «АВА ЛАБ»

Кейс 1: Анализ данных продаж для BI-системы

Описание: Загрузить данные о продажах из CSV/Excel файлов. Проверить качество данных, устранить противоречия в наименованиях товаров/категорий, подготовить данные для построения отчетов.

Цель: Подготовить данные для последующей визуализации и анализа в BI-системе.

Ожидаемый результат: Очищенный и структурированный dataset с продажами, готовый к загрузке в ClickHouse или другую аналитическую систему.

Кейс 2: Обработка данных строительных объектов

Описание: Имеются данные по стройкам: сроки этапов, площади, типы объектов. Необходимо унифицировать форматы данных, проверить на соответствие нормам, выявить аномалии в сроках выполнения работ.

Цель: Провести анализ и очистку данных строительных проектов.

Ожидаемый результат: Dataset с нормализованными данными и выделенными потенциальными проблемными объектами.

Кейс 3: Подготовка данных по клиентской базе недвижимости

Описание: Работа с базой данных клиентов: унификация ФИО, телефонных номеров, адресов. Выявление дубликатов, заполнение пропущенных контактных данных на основе имеющихся паттернов.

Цель: Научиться работать с "грязными" клиентскими данными и приводить их к единому стандарту.

Ожидаемый результат: Очищенная и дедублицированная клиентская база данных.

Для успешного выполнения кейсов рекомендуется:

1. Освоить базовые операции библиотеки Pandas для манипуляций с данными
2. Научиться работать с различными форматами данных (CSV, Excel, JSON)
3. Освоить методы обработки пропущенных и аномальных значений
4. Приобрести навыки визуального анализа данных с помощью Matplotlib/Seaborn
5. Изучить методы группировки и агрегации данных для получения сводной информации
6. Освоить базовые приемы работы с текстовыми данными в Pandas

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

7.1 Перечень информационно-коммуникационных технологий

- Проверка домашних заданий и консультирование посредством электронной почты.
- Использование электронных презентаций при проведении лекционных занятий
- Система MOODLE
- Проверка домашних заданий и консультирование посредством ЭОИС КубГУ

7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

Основное ПО
Python 3.x
NumPy
SciPy
Matplotlib
TensorFlow
Jupyter Notebook
VS Code
Google Colab
OpenOffice

8. Материально-техническое обеспечение по дисциплине (модулю)

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
1.	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
2.	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением

3.	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
4.	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.