Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Костровец Лариса Борисовна
Должность: директор
Дата подписания: 02.10.2026 00:12:42
Уникальный программный ключ:
6882606104c36dbde41c4ab93a65382136a292d6

Приложение 4
к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.О.02.ДВ.02.02 Введение в анализ данных
(индекс, наименование дисциплины в соответствии с учебным планом)

09.03.03 Прикладная информатика
(код, наименование направления подготовки/специальности)

Прикладная информатика в управлении корпоративными информационными системами
(наименование образовательной программы)

Очная форма обучения
(форма обучения)

Год набора – 2026
Донецк

Автор(ы)-составитель(и) РПД:

Лебезова Элла Михайловна, старший преподаватель кафедры информационных технологий

Заведующий кафедрой:

Брадул Наталья Валерьевна, канд. физ.-мат. наук, заведующий кафедрой информационных технологий

Рабочая программа дисциплины Б1.О.02.ДВ.02.02 Введение в анализ данных одобрена на заседании кафедры информационных технологий администрирования факультета государственной службы и управления Донецкого филиала РАНХиГС.

Протокол № 7 от «05 » марта 2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

2. Объем и место дисциплины в структуре образовательной программы

3. Содержание и структура дисциплины

4. Типы оценочных материалов, показатели и критерии их оценивания

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

7. Методические материалы по освоению дисциплины

8. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет»

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Дисциплина Б1.О.02.ДВ.02.02 Введение в анализ данных обеспечивает формирование у обучающихся следующих общепрофессиональных компетенций*:

ОТФ /ТФ и реквизиты ПС *(при наличии)* **	**Код компетенции ****	**Наименование Компетенции ****	**Код индикатора достижения компетенций ****	**Наименование индикатора достижения компетенций ****	**Образовательный результат ****
-	ОПК-2.	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.3.	Понимает принципы работы современных информационных технологий и программных средств, использует высокоуровневый язык программирования общего назначения Python для решения задач профессиональной деятельности	**Знает** синтаксис и основные библиотеки Python для анализа данных. **Умеет** использовать Python для обработки и анализа данных.

** Дисциплина может формировать компетенцию полностью или частично.*
***Должно соответствовать Приложению 1 к образовательной программе*

2. Объем и место дисциплины в структуре образовательной программы

Общий объем дисциплины:

2,00 з.е., 72 ак.час

Контактная работа обучающихся с преподавателем по видам учебных занятий: 40 ак. час на контактную работу с преподавателем, из них 18 ак. час на лекции и 18 ак. час на практические занятия. 32 ак. час на самостоятельную работу обучающихся.

Б1.О.02.ДВ.02.02. Введение в анализ данных реализуется на 3-м семестре 2-го курса после изучения дисциплин:

- Информатика и программирование
- Программирование на Python.

.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очнаяформа обучения

№ п/п	Наименование тем и (или) разделов	ВСЕГО	Объем дисциплины, ак.час												Форма текущего контроля успеваемости, промежуточной аттестации
			Контактная работа обучающихся с преподавателем по видам учебных занятий									Самостоятельная работа			
			Период теоретического обучения						Период промежуточной аттестации (сессия)						
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Каттэк	Контроль	СРкр	СРэк	СР	
			Л	ВЛ	ЛР	ПЗ									
РАЗДЕЛ 1. Основы статистики для анализа данных															
Тема 1	Введение в анализ данных	8	2	0	0	2	0	0	0	0	0	0	0	4	Контрольные вопросы, практические занятия, КР1
Тема 2	Базовые понятия описательной статистики для анализа данных	10	2	0	0	2	0	0	0	0	0	0	0	6	Контрольные вопросы, практические занятия, КР1

Тема 3	Обработка данных с использованием библиотек Python	12	4	0	0	4	0	0	0	0	0	0	0	4	Контрольные вопросы, практические занятия, КР 1
Тема 4	Основы разведочного анализа данных (EDA)	8	2	0	0	2	0	0	0	0	0	0	0	4	Контрольные вопросы, практические занятия, КР 1
РАЗДЕЛ 2. Жизненный цикл обработки данных для обучения моделей															
Тема 5	Автоматизация анализа данных средствами Python	10	2	0	0	2	0	0	0	0	0	0	0	6	Контрольные вопросы, практические занятия, КР 2
Тема 6	Подготовка данных для обучения моделей	10	4	0	0	4	0	0	0	0	0	0	0	2	Контрольные вопросы, практические занятия, КР 2
Тема 7	Создание аналитических отчётов по данным	10	2	0	0	2	0	0	0	0	0	0	0	6	Контрольные вопросы, практические занятия, КР 2
Промежуточная аттестация		4	0	0	0	0	0	0	0	0	4	0	0	0	Зачет
Итого		**72**	**18**	**0**	**0**	**18**	**0**	**0**	**0**	**0**	**4**	**0**	**0**	**32**	

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).
ИК – индивидуальные консультации.
КСР – контроль самостоятельной работы
КЭ – консультации перед экзаменом
Каттэк – контактная работа на аттестацию в период экзаменационных сессий
Контроль - контактная работа на аттестацию в период экзаменационных сессий для заочной формы обучения
СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.
СРэк – самостоятельная работа на подготовку к экзамену.
СР – самостоятельная работа в семестре на подготовку к учебным занятиям.

3.2. Содержание дисциплины

РАЗДЕЛ 1. ОСНОВЫ СТАТИСТИКИ ДЛЯ АНАЛИЗА ДАННЫХ

Тема 1. Введение в анализ данных. ОПК-2.3.

Содержание: роль анализа данных в современных ИТ-системах; жизненный цикл работы с данными; экосистема Python для анализа данных; инструмент для обработки данных и исследований в области искусственного интеллекта Google Colab; обзор библиотек NumPy, Pandas, Matplotlib, Seaborn.

Практика: настройка окружения; работа в Google Colab; загрузка CSV-файлов; первичный анализ датасета.

Тема 2. Базовые понятия описательной статистики для анализа данных. ОПК-2.3.

Содержание: меры центральной тенденции: среднее значение, медиана, мода; меры изменчивости: размах, дисперсия, стандартное отклонение; генеральная совокупность и выборка; оценивание генеральной совокупности; распределение вероятностей: нормальное распределение, равномерное распределение, геометрическое распределение.

Практика: вычисление мер центральной тенденции в Python; построение гистограмм распределения в Python.

Тема 3. Обработка данных с использованием библиотек Python. ОПК-2.3.

Содержание: библиотеки NumPy и Pandas, массивы NumPy и операции над ними; структура DataFrame; загрузка и сохранение данных; фильтрация, сортировка и группировка данных; работа с пропущенными значениями; работа с дубликатами; создание новых признаков.

Практика: анализ наборов данных; обработка пропусков; группировка и агрегация; преобразование категориальных признаков.

Тема 4. Основы разведочного анализа данных (EDA). ОПК-2.3.

Содержание: основные подходы EDA; корреляции и зависимости; основные типы визуализаций; распределения и выбросы; библиотеки Matplotlib и Seaborn; интерпретация графиков.

Практика: однофакторный и многофакторный анализ данных с помощью построения histogram, boxplot, scatterplot; анализ зависимостей и heatmap; визуальный поиск закономерностей и выбросов.

РАЗДЕЛ 2. ЖИЗНЕННЫЙ ЦИКЛ ОБРАБОТКИ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ

Тема 5. Автоматизация анализа данных средствами Python. ОПК-2.3.

Содержание: автоматизированный EDA; библиотека ydata_profiling; генерация аналитических отчётов; анализ структуры и качества данных; обзор дополнительных библиотек: missingno, sweetviz, pandas-profiling (исторически),

plotly express.

Практика: генерация автоматического отчёта; анализ пропусков и распределений; интерпретация результатов profiling-отчётов.

Тема 6. Подготовка данных для обучения моделей. ОПК-2.3.

Содержание: понятие признака; кодирование категориальных данных; масштабирование и нормализация; работа с выбросами; разделение данных на обучающую и тестовую выборки; понятие quality data.

Практика: подготовка датасета к анализу; кодирование признаков; нормализация; подготовка данных для обучения модели.

Тема 7. Создание аналитических отчётов по данным. ОПК-2.3.

Содержание: постановка аналитической задачи; выбор признаков; визуальный анализ; формулировка выводов; подготовка аналитического отчёта.

Практика: выполнение мини-проекта на реальном датасете; проведение EDA; визуализация данных; защита результатов анализа.

4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.О.02.ДВ.02.02. Введение в анализ данных входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам (модулям) образовательной программы составляет фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых

каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа – это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	Ответ считается верным, если правильно указаны цифры или буквы

Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	Прочитайте текст, выберите правильные ответы	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).	Ответ считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого)
Задание закрытого типа на установление последовательнос ти	Прочитайте текст и установите последовательность	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	Ответ считается верным, если правильно указана вся последовательность цифр

Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа
Задание открытого типа с развернутым ответом	Прочитайте текст и запишите развернутый обоснованный ответ	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ	Ответ считается верным: 1. Отсутствие фактических ошибок. 2. Раскрытие объема используемых понятий (полнота ответа). 3. Обоснованность ответа (наличие аргументов). 4. Логическая последовательность излагаемого материала.

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Оценка по шкале ECTS	**Сумма баллов за все виды учебной деятельности**	**Оценка по государственной шкале**	**Определение**
A	90 – 100	«Отлично»	отличное выполнение с незначительным количеством неточностей
B	80 – 89	«Хорошо»	в целом правильно выполненная работа с незначительным количеством ошибок (до 10%)
C	75 – 79		в целом правильно выполненная работа с незначительным количеством ошибок (до 15%)
D	70 – 74	«Удовлетворительно»	неплохо, но со значительным количеством недостатков
E	60 – 69		выполнение удовлетворяет минимальные критерии
FX	35 – 59	«Не удовлетворительно»	с возможностью повторной сдачи
F	0 – 34		с обязательным повторным изучением дисциплины (выставляется комиссией)

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
100 баллов	100 баллов	100 баллов	100 баллов

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины Б1.О.02.ДВ.02.02 Введение в анализ данных используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Контрольные вопросы для проведения опроса, задания открытого типа

на практических занятиях, контрольные задания

Таблица 5.1.

Распределение баллов по видам учебной деятельности (БРС)

Раздел/Темы	Формы текущего контроля		КЗР
	УО	ПЗ	
Р-1. / Т-1	3	7	15
Р-1. / Т-2	3	7	
Р-1. / Т-3	3	7	
Р-1. / Т-4	3	7	15
Р-2. / Т-5	3	7	
Р-2. / Т-6	3	7	
Р-2. / Т-7	3	7	
Итого: 100 б	18	7	30

УО – устный опрос;
ТЗ – тестовое задание;
КЗ – контрольные задания;
ПЗ – практическое занятие;
Д – доклад;
КЗР – контрольные работы по разделу.

Критерии оценивания опроса:

Балы	Описание критерия
3	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
2	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.
1	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
0	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

0* - в журнал академической группы не выставляется

Критерии оценивания практических занятий:

Балы	Описание критерия	
3	Свыше 90% правильных ответов.	Обучающийся демонстрирует глубокое познание в освоенном материале.
2	Свыше 70% правильных ответов.	Обучающимся материал освоен полностью, без существенных ошибок.

1	Реализовано более 50% поставленных задач	Обучающимся материал освоен не полностью, имеются значительные пробелы в знаниях.
0	Реализовано менее 30% поставленных задач.	Обучающимся материал не освоен, знания обучающегося ниже базового уровня.

0* - в журнал академической группы не выставляется

Критерии оценивания контрольных заданий:

Балы	Описание критерия
12-15	Обучающимся задание выполнено без ошибок и в полном объеме.
8-11	Обучающимся в целом задание выполнено, имеются отдельные неточности или недостаточно полные ответы, не содержащие ошибок.
5-7	Обучающимся допущены отдельные ошибки при выполнении задания
0-4	У обучающегося отсутствуют ответы на большинство вопросов задачи, задание не выполнено или выполнено не верно.

0* - в журнал академической группы не выставляется

5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных работ):

РАЗДЕЛ 1. ОСНОВЫ СТАТИСТИКИ ДЛЯ АНАЛИЗА ДАННЫХ

Тема 1. Введение в анализ данных.

Контрольные вопросы для проведения опроса:

1. Что понимается под аналитикой данных?
2. Перечислите основные этапы работы с данными.
3. Для чего используется Jupyter Notebook?
4. Какие библиотеки Python наиболее часто применяются для анализа данных?
5. В чём различие между библиотеками NumPy и Pandas?
6. Что такое датасет и какие форматы хранения данных наиболее распространены?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «penguins.csv»:

- загрузить датасет средствами Pandas;
- вывести первые 5 строк;
- определить количество строк и столбцов;
- определить типы признаков;
- выделить категориальные и числовые признаки;
- сформулировать краткий вывод о структуре датасета.

Тема 2. Базовые понятия описательной статистики для анализа

данных.

Контрольные вопросы для проведения опроса:

1. Перечислите меры центральной тенденции. В чём их отличие?
2. Какие меры изменчивости используют при анализе данных?
3. Какая мера центральной тенденции, применима к категориальным данным?
4. Перечислите преимущества метода межквартильного размаха?
5. Когда можно применить правило трёх сигм?
6. Перечислите стратегии обработки пропущенных данных?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «Population.csv»:

- вычислить среднее значение населения Болгарии за все годы, представленные в таблице. Ответ округлить до целого числа;
- вычислить среднее значение населения Болгарии за пять последних лет, представленных в таблице. Верно ли, что в среднем за пять последних лет наблюдается прирост населения по сравнению со средним значением за все годы?
- вычислить медиану населения Болгарии за все годы из таблицы. Ответ округлите до целого числа.
- найти страну, население которой ведет себя похожим образом (как в Болгарии) - сначала растет, а потом падает. Что это за страна (из перечисленных)?
- вычислить несмещенное выборочное стандартное отклонение для населения страны, найденной в предыдущем задании. Ответ округлите до целого числа.
- вычислить несмещенное выборочное стандартное отклонение для населения Болгарии. Верно ли, что в Болгарии "разброс" относительно среднего количества населения больше, чем в стране, найденной в предыдущем задании?
-

Тема 3. Обработка данных с использованием библиотек Python.

Контрольные вопросы для проведения опроса:

7. Что представляет собой DataFrame?
8. Какие способы обращения к строкам и столбцам DataFrame существуют?
9. Для чего используются фильтрация и группировка данных?
10. Что такое пропущенные значения и почему они возникают?
11. Перечислите основные способы обработки пропусков.
12. Что такое агрегация данных?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «penguins.csv»:

- определить количество пропущенных значений в каждом столбце;
- удалить или заполнить пропущенные значения;
- отфильтровать пингвинов вида Adelie с массой более 3000 г;
- отсортировать датасет по массе тела;
- определить самых тяжёлых пингвинов;
- создать новый признак body_mass_kg.

Тема 4. Основы разведочного анализа данных (EDA).

Контрольные вопросы для проведения опроса:

1. Что понимается под разведочным анализом данных (EDA)?
2. Для чего используется визуализация данных?
3. В каких случаях применяется histogram?
4. Что показывает boxplot?
5. Что позволяет выявить scatter plot?
6. Что такое корреляция признаков?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «penguins.csv»:

- построить histogram распределения массы тела;
- построить boxplot массы по видам;
- построить scatter plot зависимости массы тела от длины ласт;
- построить heatmap корреляций;
- проанализировать полученные визуализации;
- сформулировать выводы о зависимостях между признаками.

РАЗДЕЛ 2. ЖИЗНЕННЫЙ ЦИКЛ ОБРАБОТКИ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ

Тема 5. Автоматизация анализа данных средствами Python.

Контрольные вопросы для проведения опроса:

1. Что такое автоматизированный разведочный анализ данных?
2. Для чего используется библиотека ydata_profiling?
3. Какие сведения содержит profiling-отчёт?
4. Что позволяет выявить анализ пропусков?
5. Какие библиотеки Python используются для автоматизированной визуализации данных?
6. В чём преимущества автоматизированного анализа данных?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «penguins.csv»:

- сформировать profiling-отчёт средствами ydata_profiling;
- определить признаки с пропущенными значениями;
- выявить признаки с высокой корреляцией;
- определить возможные выбросы;
- проанализировать распределения признаков;
- подготовить краткий аналитический вывод по отчёту.

Тема 6. Подготовка данных для обучения моделей.

Контрольные вопросы для проведения опроса:

1. Что такое признак (feature) в машинном обучении?
2. Для чего требуется кодирование категориальных данных?
3. В чём различие между нормализацией и стандартизацией?
4. Что такое обучающая и тестовая выборка?
5. Почему качество данных влияет на качество модели?
6. Какие проблемы могут возникать при наличии выбросов?

Практические занятия:

Задание открытого типа с развернутым ответом.

Для датасета «penguins.csv»:

- выполнить кодирование категориальных признаков;
- выполнить масштабирование числовых признаков;
- разделить данные на обучающую и тестовую выборки;
- определить признаки, наиболее подходящие для определения вида пингвина;
- объяснить выбор признаков;
- подготовить датасет для дальнейшего применения моделей машинного обучения.

Тема 7. Создание аналитических отчётов по данным.

Контрольные вопросы для проведения опроса:

1. Какие этапы включает аналитический проект?
2. Что такое аналитический вывод?
3. Почему важно визуально исследовать данные перед построением модели?
4. Какие признаки можно считать наиболее информативными?
5. Что такое интерпретация результатов анализа?
6. Какие ошибки наиболее часто допускаются при анализе данных?

Практические занятия:

Задание открытого типа с развернутым ответом.

На основе предложенного датасета:

- выполнить разведочный анализ данных;
- построить не менее трёх визуализаций;
- выявить зависимости между признаками;
- определить наиболее информативные признаки;
- сформулировать аналитические выводы;
- подготовить краткий отчёт с интерпретацией результатов.

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной работой по разделу (далее – КР). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КР в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КР составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КР в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной работы	Максимальное количество баллов за работу в рамках КР, которое может набрать студент	Коэффициент веса контрольной работы	Результат контрольной работы, участвующий в формировании итоговой балльной оценки по дисциплине
КР 1	100	0,15	15
КР 2	100	0,15	15
Итого:	x	0,30	30

Формула расчета результата контрольной работы:

Результат контрольной работы = Количество баллов за работу в рамках КР X Коэффициент веса контрольной работы.

5.4. Формы текущего контроля успеваемости обучающихся в рамках КР и типовые оценочные материалы:

КР-1

Раздел 1. Основы статистики для анализа данных.

Задание 1. Загрузка и первичный анализ данных

Для датасета «penguins.csv»:

- загрузите датасет средствами Pandas;
- выведите первые 5 строк;
- определите количество строк и столбцов;
- определите типы признаков;
- выделите категориальные и числовые признаки;
- определите наличие пропущенных значений.

Сформулируйте краткий вывод о структуре датасета.

Задание 2. Обработка данных средствами Pandas

Для датасета «penguins.csv»:

- удалите строки с пропущенными значениями признака sex;
- отберите пингвинов вида Adelie с массой более 3000 г;
- отсортируйте полученный датасет по массе тела;
- определите трёх самых тяжёлых пингвинов;
- создайте новый признак body_mass_kg.

Поясните, зачем выполняется фильтрация и создание новых признаков.

Задание 3. Группировка и агрегация данных

Для датасета «penguins.csv»:

- сгруппируйте данные по виду пингвинов (species);
- найдите: среднюю массу тела, среднюю длину клюва, максимальную длину ласт;
- определите, какой вид является самым крупным в среднем.

Сделайте вывод по результатам группировки.

Задание 4. Разведочный анализ данных (EDA)

Для датасета «penguins.csv»:

- постройте histogram распределения массы тела;
- постройте boxplot массы тела по видам;
- постройте scatter plot зависимости массы тела от длины ласт;
- проанализируйте графики;
- определите: наличие выбросов, возможные зависимости, различия между видами.

Сформулируйте аналитический вывод.

КР-2

Раздел 2. Жизненный цикл обработки данных для обучения моделей.

Задание 1. Автоматизированный анализ данных

Для датасета «penguins.csv»:

- сформируйте profiling-отчёт средствами ydata_profiling;
- определите:
- признаки с пропущенными значениями,
- признаки с высокой корреляцией,
- признаки с наибольшим разбросом значений;
- выявите возможные выбросы;
- сформулируйте краткий аналитический отчёт по результатам анализа.

–

Задание 2. Подготовка данных для машинного обучения

Для датасета «penguins.csv»:

– выполните кодирование категориальных признаков;
– выполните масштабирование числовых признаков;
– разделите данные на обучающую и тестовую выборки;
– определите признаки, наиболее полезные для определения вида пингвина;
– объясните свой выбор.

Сформулируйте вывод о качестве подготовленных данных.

Задание 3. Анализ зависимостей между признаками

Для датасета «penguins.csv»:

– постройте heatmap корреляций;
– определите признаки с наиболее сильной взаимосвязью;
– постройте pairplot для основных числовых признаков;
– определите, какие признаки лучше всего разделяют виды пингвинов.

Сделайте вывод о пригодности признаков для задач машинного обучения.

Задание 4. Мини-проект по анализу данных

На основе предложенного датасета:

– выполните разведочный анализ данных;
– постройте не менее трёх визуализаций;
– выявите закономерности и зависимости;
– определите наиболее информативные признаки;
– предложите, какие задачи машинного обучения можно решать на данном датасете;
– подготовьте краткий аналитический отчёт.

В отчёте необходимо:

описать структуру данных;

представить результаты анализа;

сделать итоговые выводы.

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация по дисциплине «Введение в анализ данных» проводится в форме зачёта в третьем семестре в письменной форме. Обучающийся получает три теоретических вопроса и одно практическое задание.

Теоретические вопросы направлены на проверку:

– понимания основных методов анализа данных;

– знания библиотек Python для обработки и визуализации данных;

– понимания принципов разведочного анализа данных и подготовки данных для задач машинного обучения.

Практическое задание направлено на проверку умений:

– загружать и обрабатывать данные;

– выполнять базовый анализ данных;

– строить визуализации;

интерпретировать результаты анализа.

6.2. Типовые оценочные материалы промежуточной аттестации.

Вопросы к зачету:

1. Что понимается под аналитикой данных?
2. Какие этапы включает процесс анализа данных?
3. Для чего используется библиотека NumPy?
4. Что представляет собой DataFrame в библиотеке Pandas?
5. Какие основные операции выполняются над DataFrame?
6. Что такое фильтрация данных?
7. Для чего применяется группировка данных?
8. Что такое пропущенные значения и какие способы их обработки существуют?
9. Что такое разведочный анализ данных (EDA)?
10. Для чего используется визуализация данных?
11. Какие задачи решают histogram и boxplot?
12. Что позволяет выявить scatter plot?
13. Что такое корреляция признаков?
14. Для чего используется heatmap?
15. Какие библиотеки Python применяются для визуализации данных?
16. Что такое автоматизированный анализ данных?
17. Для чего используется библиотека ydata_profiling?
18. Какие сведения можно получить из profiling-отчёта?

19. Что такое признак (feature) в машинном обучении?
20. Для чего требуется кодирование категориальных признаков?
21. В чём различие между нормализацией и стандартизацией?
22. Что такое обучающая и тестовая выборки?
23. Почему качество данных влияет на качество модели?
24. Какие признаки можно считать информативными?
25. Какие задачи машинного обучения можно решать после подготовки данных?

Пример практического задания

Для датасета «penguins.csv»:

1. Загрузить датасет средствами Pandas.
2. Определить количество строк и столбцов.
3. Определить наличие пропущенных значений.
4. Удалить строки с пропущенными значениями признака sex.
5. Отфильтровать пингвинов вида Adelie с массой более 3000 г.
6. Создать новый признак body_mass_kg.
7. Построить:
   - histogram распределения массы тела;
   - scatter plot зависимости массы тела от длины ласт.
8. Определить:
   - наличие выбросов;
   - зависимость между признаками;
   - различия между видами пингвинов.
9. Сформулировать краткий аналитический вывод по результатам анализа.

6.3. Критерии и шкала оценивания на основе БРС.

Соответствие государственной шкалы оценивания академической успеваемости и шкалы ECTS при зачёте

Оценка по шкале ECTS	**Сумма баллов за все виды учебной деятельности**	**Оценка по государственной шкале**	**Определение**
A	90 – 100	«Отлично»	отличное выполнение с незначительным количеством неточностей
B	80 – 89	«Хорошо»	в целом правильно выполненная работа с незначительным количеством ошибок (до 10%)
C	75 – 79		в целом правильно выполненная работа с незначительным количеством ошибок (до 15%)
D	70 – 74	«Удовлетворительно»	неплохо, но со значительным количеством недостатков
E	60 – 69		выполнение удовлетворяет минимальные критерии
FX	35 – 59	«Не удовлетворительно»	с возможностью повторной сдачи
F	0 – 34		с обязательным повторным изучением дисциплины (выставляется комиссией)

6.4. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий

Компьютер с операционной системой RedOS или Windows с устойчивым Интернет-соединением для работы с ноутбуками Google Colab, программные продукты с открытой лицензией: PyCharm Community Edition, Visual Studio Code, Pandas, Scikit-learn, Anaconda, Jupyter Notebook.

7. Методические материалы по освоению дисциплины

Получение углубленных знаний по изучаемой дисциплине достигается за счет дополнительных часов к аудиторной работе самостоятельной работы студентов. Выделяемые часы целесообразно использовать для знакомства с дополнительной научной литературой по проблематике дисциплины, анализа научных концепций и современных подходов к осмыслению рассматриваемых проблем. К самостоятельному виду работы студентов относится работа в библиотеках, в электронных поисковых системах и т.п. по сбору материалов, необходимых для проведения практических занятий или выполнения конкретных заданий преподавателя по изучаемым темам. Студенты могут установить диалог с преподавателем, получать консультации по выполнению заданий. В качестве оценочных средств на протяжении семестра используются практические задания.

Обучение по дисциплине «Введение в анализ данных» предполагает изучение курса на аудиторных занятиях (лекции, практические занятия) и самостоятельную работу студентов. Практические занятия дисциплины предполагают их проведение в различных формах с целью выявления полученных знаний, умений, навыков и компетенций с проведением контрольных мероприятий. С целью обеспечения успешного обучения студент должен готовиться к лекции, поскольку она является важнейшей формой организации учебного процесса, поскольку:

– знакомит с новым учебным материалом;

– разъясняет учебные элементы, трудные для понимания;

– систематизирует учебный материал;

– ориентирует в учебном процессе.

Работа обучающегося на лекции:

Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность обучающегося. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим обучающимся.

Подготовка к практическим занятиям:

Подготовку к каждому практическому занятию каждый обучающийся должен начать с ознакомления с планом, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции. Все новые понятия по изучаемой теме необходимо внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности обучающегося свободно ответить на теоретические вопросы практического занятия, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Структура практического занятия:

В зависимости от содержания и количества отведенного времени на изучение каждой темы может практическое занятие состоять из четырех-пяти частей:

1. Устный опрос.

2. Обсуждение теоретических вопросов, определенных программой дисциплины.

3. Выполнение практических заданий с последующим разбором полученных результатов или обсуждение практического задания, выполненного дома.

4. Подведение итогов занятия.

Работа с литературными источниками:

В процессе подготовки к практическим занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем, что позволяет обучающимся проявить свою индивидуальность в рамках выступления на занятиях, выявить широкий спектр мнений по изучаемой проблеме.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Панов, М. А. Анализ данных с использованием языка программирования Python : учебное пособие / М. А. Панов. — Екатеринбург : УрГЭУ, 2024. — 329 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/481577 (дата обращения: 09.05.2026). — Режим доступа: для авториз. пользователей.

2. Гущина, О. М. Анализ и визуализация данных : учебно-методическое пособие / О. М. Гущина, О. В. Аникина, Е. В. Желнина. — Тольятти : ТГУ, 2025. — 204 с. — ISBN 978-5-8259-1712-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/517526 (дата обращения: 09.05.2026). — Режим доступа: для авториз. пользователей.

8.2. Дополнительная литература

3. Никульчев, Е. В. Системы сбора и предобработки данных. Методы статистического анализа с использованием Google Colab : учебное пособие / Е. В. Никульчев, А. С. Алексеенко, Д. Ю. Ильин. — Москва : РТУ МИРЭА, 2023. — 121 с. — ISBN 978-5-7339-1948-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/382739 (дата обращения: 09.05.2026). — Режим доступа: для авториз. пользователей.

8.3.Нормативные правовые документы и иная правовая информация

1. Конституция Российской Федерации. – Текст : электронный // Сайт Президента Российской Федерации. – URL: http://www.kremlin.ru/acts/constitution

8.4 Интернет-ресурсы

1. Информационно-правовой портал ГАРАНТ.РУ. – URL: https://www.garant.ru/

2. Научная электронная библиотека eLIBRARY.RU. – URL: https://elibrary.ru/

3. Научная электронная библиотека «КиберЛенинка». – URL: https://cyberleninka.ru

4. Электронно-библиотечная система «Лань». – URL: http://e.lanbook.com

5. База знаний по Pandas – URL: https://pandas.pydata.org/docs/
6. База знаний по Seaborn– URL: https://seaborn.pydata.org/
7. Документация по Python – URL: https:// https://docs.python.org/3/

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Материально-техническое обеспечение дисциплины включает в себя:

- лекционные аудитории, оборудованные видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения, экраном;
- помещения для проведения практических занятий, оборудованные учебной мебелью.

Дисциплина поддержана соответствующими программными продуктами с открытой лицензией: PyCharm Community Edition, Visual Studio Code, Pandas, Scikit-learn, Anaconda, Jupyter Notebook.

Вуз обеспечивает каждого обучающегося рабочим местом в компьютерном классе в соответствии с объемом изучаемых дисциплин, обеспечивает выход в сеть Интернет.

Помещения для самостоятельной работы обучающихся включают следующую оснащенность: столы аудиторные, стулья, доски аудиторные, компьютеры с подключением к локальной сети института (для компьютерных аудиторий) и Интернет. Для изучения учебной дисциплины используются автоматизированная библиотечная информационная система и электронные библиотечные системы.