

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение высшего  
образования  
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 41

УТВЕРЖДАЮ

Руководитель образовательной программы

доц., к.т.н. \_\_\_\_\_

(должность, уч. степень, звание)

Е.Л. Турнецкая \_\_\_\_\_

(инициалы, фамилия)



(подпись)

«27» июня 2024 г.

Лист согласования рабочей программы дисциплины

Программу составил (а)

Проф., д-р техн. наук,  
проф. \_\_\_\_\_

(должность, уч. степень, звание)

  
(подпись, дата)

Т.М. Тагарникова \_\_\_\_\_  
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 41

«18» июня 2024 г, протокол № 11-2023/24

Заведующий кафедрой № 41



д.т.н., проф. \_\_\_\_\_

(уч. степень, звание)

  
(подпись, дата)

Г.А. Коржавин \_\_\_\_\_  
(инициалы, фамилия)

Заместитель директора института №4 по методической работе

доц., к.т.н. \_\_\_\_\_

(должность, уч. степень, звание)

  
(подпись, дата)

А.А. Фоменкова \_\_\_\_\_  
(инициалы, фамилия)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Введение в анализ данных»  
(Наименование дисциплины)

Код направления подготовки/ специальности	09.03.03
Наименование направления подготовки/ специальности	Прикладная информатика
Наименование направленности	Прикладная информатика в информационной сфере
Форма обучения	заочная
Год приема	2024

## Аннотация

Дисциплина «Введение в анализ данных» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/ специальности 09.03.03 «Прикладная информатика» направленности «Прикладная информатика в информационной сфере». Дисциплина реализуется кафедрой «№41».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-1 «Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач»

УК-2 «Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений»

ПК-13 «способность применять системный подход и математические методы в формализации решения прикладных задач»

ПК-14 «Способность использовать знание основных методов искусственного интеллекта в последующей профессиональной деятельности в качестве научных сотрудников, преподавателей образовательных организаций высшего образования, инженеров, технологов»

Содержание дисциплины охватывает круг вопросов, связанных с изучением и применением методов анализа данных и извлечения информации, закономерностей, знаний и шаблонов из больших объемов необработанных данных.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные работы, практические занятия, самостоятельная работа обучающегося.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме дифференцированного зачета.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Язык обучения по дисциплине «русский»

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины: формирование компетенций, позволяющих использовать соответствующий математический аппарат и инструментальные средства для анализа данных и извлечения информации, закономерностей, знаний и шаблонов из больших объемов необработанных данных.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Универсальные компетенции	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.3.1 знать методики поиска, сбора и обработки информации, в том числе с использованием информационных технологий, включая интеллектуальные УК-1.У.3 уметь оценивать информацию на достоверность; сохранять и передавать данные с использованием цифровых средств УК-1.В.1 владеть навыками критического анализа и синтеза информации, в том числе с помощью цифровых инструментов
Универсальные компетенции	УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.3.3 знать возможности и ограничения применения цифровых инструментов для решения поставленных задач УК-2.У.1 уметь проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения УК-2.В.3 владеть навыками использования цифровых средств для решения поставленной задачи
Профессиональные компетенции	ПК-13 способность применять системный подход и математические методы в формализации решения прикладных задач	ПК-13.У.1 уметь осуществлять формализацию задач исследования информационных процессов и систем ПК-13.В.1 владеть навыками решения задач анализа информационных процессов и систем
Профессиональные компетенции	ПК-14 Способность использовать знание основных методов искусственного	ПК-14.У.1 уметь применять методы машинного обучения, подготовки данных и интерпретации результатов

	интеллекта в последующей профессиональной деятельности в качестве научных сотрудников, преподавателей образовательных организаций высшего образования, инженеров, технологов	
--	--	--

## 2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- Теория вероятностей,
- Математика. Математический анализ
- Теория систем и системный анализ,
- Информационные системы и технологии.

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и используются при изучении других дисциплин:

- Интеллектуальные информационные системы,
- Машинное обучение
- Информационная безопасность.

## 3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№6
1	2	3
<b>Общая трудоемкость дисциплины, ЗЕ/ (час)</b>	3/ 108	3/ 108
<b>Из них часов практической подготовки</b>	4	4
<b>Аудиторные занятия, всего час.</b>	12	12
в том числе:		
лекции (Л), (час)	4	4
практические/семинарские занятия (ПЗ), (час)	4	4
лабораторные работы (ЛР), (час)	4	4
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
<b>Самостоятельная работа, всего (час)</b>	96	96
<b>Вид промежуточной аттестации:</b> зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Дифф. Зач.	Дифф. Зач.

Примечание: \*\* кандидатский экзамен

## 4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ) (час)	ЛР (час)	КП (час)	СРС (час)
Семестр 6					
Раздел 1. Введение Тема 1.1. Данные и их измерение. Тема 1.2. Шкалы Тема 1.3. Характеристика и классификация задач анализа данных Тема 1.4. Этапы решения задачи анализа данных.	0,5	2			12
Раздел 2. Статистический анализ данных Тема 2.1. Дескриптивный анализ Тема 2.2. Анализ связей Тема 2.3. Многомерный статистический анализ Тема 2.4. Анализ временных рядов	1				18
Раздел 3. Кластерный анализ Тема 3.1. Постановка задачи кластерного анализа Тема 3.2. Метрики расстояний Тема 3.3. Методы кластерного анализа	1		2		25
Раздел 4. Классификация Тема 4.1. Постановка задачи классификации Тема 4.2. Методы классификации	1	2	2		25
Раздел 5. Заключение Тема 5.1. Перспективы развития анализа данных.	0,5				16
Итого в семестре:	4	4	4		96
Итого	4	4	4	0	96

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

#### 4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
Раздел 1. Введение	<p><b>Тема 1.1. Данные и их измерение.</b> Понятие данных. Измерения. Типы наборов данных. Форматы хранения данных. Классификация видов данных. Постановка задачи анализа данных. История развития технологий анализа данных.</p> <p><b>Тема 1.2. Шкалы</b> Типы шкал измерений. Номинальная, порядковая, интервальная, относительная и дихотомическая шкалы.</p> <p><b>Тема 1.3. Характеристика и классификация задач анализа данных</b> Статистические методы: дескриптивный анализ и описание исходных данных; анализ связей; многомерный статистический</p>

	<p>анализ; анализ временных рядов.  Методы машинного обучения: искусственные нейронные сети; деревья решений, машина опорных векторов, метод ближайшего соседа, метод Байеса.  <b>Тема 1.4. Этапы решения задачи анализа данных.</b>  Методы сбора и подготовки исходного набора данных. Формулировка цели. Выбор метода анализа данных. Представление результатов.</p>
Раздел 2. Статистический анализ данных	<p><b>Тема 2.1. Дескриптивный анализ</b>  Проверка гипотез стационарности, нормальности, независимости, однородности, оценка вида функции распределения и ее параметров  <b>Тема 2.2. Анализ связей</b>  Корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ  <b>Тема 2.3. Многомерный статистический анализ</b>  Компонентный анализ, дискриминантный анализ, многомерный регрессионный анализ, канонические корреляции  <b>Тема 2.4. Анализ временных рядов</b>  Динамические модели и прогноз на основе временных рядов</p>
Раздел 3. Кластерный анализ	<p><b>Тема 3.1. Постановка задачи кластерного анализа</b>  Поиск существующих структур. Оценка качества кластеризации. Процесс кластеризации.  <b>Тема 3.2. Метрики расстояний</b>  Понятие метрики, мера близости объектов. Евклидово расстояние, другие метрики.  <b>Тема 3.3. Методы кластерного анализа</b>  Иерархические алгоритмы, неиерархические алгоритмы</p>
Раздел 4. Классификация сети	<p><b>Тема 4.1. Постановка задачи классификации</b>  Системное распределение изучаемых предметов по существенным признакам. Правила классификации.  <b>Тема 4.2. Методы классификации</b>  Дерево решений, случайный лес, метод опорных векторов, метод ближайшего соседа, метод Байеса, нейронные сети</p>
Раздел 5. Заключение	<p><b>Тема 5.1. Перспективы развития анализа данных.</b>  Приложение методов анализа данных к актуальным задачам. Другие задачи анализа данных</p>

#### 4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 6					
1	Анализ одномерного набора данных	Занятие по моделированию реальных условий	1	0,5	2
2	Анализ двумерного набора данных	Занятие по моделированию реальных условий	1	0,5	2
3	Кластерный	Игровое	2	1	3

	анализ: нормализация данных	проектирование			
Всего			4	2	

#### 4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 6				
1	Лабораторная работа 1: Кластеризация данных	2	1	3,5
2	Лабораторная работа 2: Классификация данных	2	1	4,5
Всего		4	2	

4.5. Курсовое проектирование/ выполнение курсовой работы  
Учебным планом не предусмотрено

#### 4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 6, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	50	50
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)	10	10
Домашнее задание (ДЗ)		
Контрольные работы заочников (КРЗ)	20	20
Подготовка к промежуточной аттестации (ПА)	16	16
Всего:	96	96

#### 5. Перечень учебно-методического обеспечения

для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7-11.

#### 6. Перечень печатных и электронных учебных изданий

Перечень печатных и электронных учебных изданий приведен в таблице 8.

Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных)
--------------	--------------------------	---

адрес		экземпляров)
004 А 51	Алпайдин, Э. Машинное обучение: новый искусственный интеллект / Э. Алпайдин. - М.: Альпина Паблишер: Точка, 2017. - 208 с.	3
004 Т 23	Татарникова, Т.М. Системы искусственного интеллекта: учебник / Т. М. Татарникова; С.-Петерб. гос. ун-т аэрокосм. приборостроения. - Санкт-Петербург: Изд-во ГУАП, 2024. - 301 с.	5
004 Т 23	Татарникова, Т.М. Анализ данных в прикладных задачах обеспечения информационной безопасности: монография / Т. М. Татарникова; С.-Петербург. гос. ун-т аэрокосм. приборостроения. - СПб.: Изд-во ГУАП, 2018. - 115 с.	5
004 Т 23	Татарникова, Т. М. Анализ данных: учебно-методическое пособие / Т. М. Татарникова; С.-Петерб. гос. ун-т аэрокосм. приборостроения. - Санкт-Петербург: Изд-во ГУАП, 2021. - 121 с.	5

#### 7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
1. <a href="http://www.machinelearning.ru">http://www.machinelearning.ru</a>	1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. <a href="https://www.coursera.org/learn/machine-learning">https://www.coursera.org/learn/machine-learning</a>	2. Machine learning by Andrew Ng,
3. <a href="https://www.coursera.org/specializations/machine-learning-data-analysis">https://www.coursera.org/specializations/machine-learning-data-analysis</a>	3. Курс «Анализ данных» от школы анализа данных,
4. <a href="https://it.mail.ru/video/playlists/16/">https://it.mail.ru/video/playlists/16/</a>	4. Методы обработки больших данных, IT mail.ru,

#### 8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	Google Colab
2	Jupyter Notebook



8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
	Не предусмотрено

## 9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Мультимедийная лекционная аудитория	
2	Класс для деловой игры	
3	Вычислительная лаборатория	

## 10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Дифференцированный зачёт	Список вопросов; Тесты; Задачи.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 –Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«отлично» «зачтено»	<ul style="list-style-type: none"> <li>– обучающийся глубоко и всесторонне усвоил программный материал;</li> <li>– уверенно, логично, последовательно и грамотно его излагает;</li> <li>– опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления;</li> <li>– умело обосновывает и аргументирует выдвигаемые им идеи;</li> <li>– делает выводы и обобщения;</li> <li>– свободно владеет системой специализированных понятий.</li> </ul>

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«хорошо» «зачтено»	<ul style="list-style-type: none"> <li>– обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы;</li> <li>– не допускает существенных неточностей;</li> <li>– увязывает усвоенные знания с практической деятельностью направления;</li> <li>– аргументирует научные положения;</li> <li>– делает выводы и обобщения;</li> <li>– владеет системой специализированных понятий.</li> </ul>
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> <li>– обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы;</li> <li>– допускает несущественные ошибки и неточности;</li> <li>– испытывает затруднения в практическом применении знаний направления;</li> <li>– слабо аргументирует научные положения;</li> <li>– затрудняется в формулировании выводов и обобщений;</li> <li>– частично владеет системой специализированных понятий.</li> </ul>
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> <li>– обучающийся не усвоил значительной части программного материала;</li> <li>– допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении;</li> <li>– испытывает трудности в практическом применении знаний;</li> <li>– не может аргументировать научные положения;</li> <li>– не формулирует выводов и обобщений.</li> </ul>

10.3. Типовые контрольные задания или иные материалы.  
Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
	Учебным планом не предусмотрено	

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. Зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
1	Назначение обучающего множества при решении задачи классификации	УК-1.3.1
2	Базовые характеристики статистического распределения.	
3	Обобщающие показатели, характеризующие типические значения набора данных.	
4	Показатели, характеризующие разброс данных.	
5	Показатели, характеризующие степень симметричности данных.	
6	Цель применения метода наименьших квадратов в регрессионном анализе.	
7	Назначение моды. Случай вычисления модального интервала.	
8	Назначение персентилия, квантиля единицы измерения.	
9	Отличие выборочного стандартного отклонения от	

	стандартного отклонения генеральной совокупности.	
10	Анализ одномерного набора данных	УК-1.У.3
11	Анализ двумерного набора данных.	
12	Корреляционный анализ.	
13	Основная задача регрессионного анализа.	
14	Интерпретация коэффициентов линейной регрессии $a_0, a_1$ .	
15	Вычисление медианы для набора данных с четным и нечетным количеством значений.	УК-1.В.1
16	Меры близости между кластерами.	УК-2.3.3
17	Метрики, применяемые в кластерном анализе.	
18	Вычисление центра кластера.	УК-2.У.1
19	Ограничения кластерного анализа.	
20	Назначение тестового множества при решении задачи классификации	УК-2.В.3
21	Методы классификации.	
22	Оценка ошибки классификации	
23	Определение кластера.	ПК-13.У.1
24	Выражение для вычисления индекса Gini при построении	
25	Требования к выбору параметра $k$ в алгоритме « $k$ -ближайшего соседа»	
26	Постановка задачи классификации	ПК-13.В.1
27	Этапы процесса классификации, их характеристики	
28	Этапы решения задачи кластерного анализа.	
29	Иерархические методы кластеризации.	ПК-14.У.1
30	Этапы построения деревьев классификации дерева классификации.	
31	Алгоритм ближайшего соседа.	
32	Алгоритм $k$ -средних.	
33	Функции активации нейронов.	
34	Нейронная сеть Кохонена.	
35	Алгоритм обучения многослойного перцептрона	

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

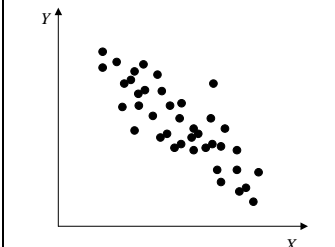
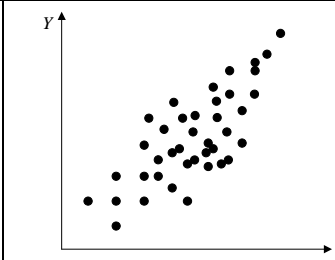
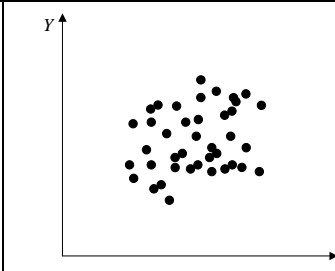
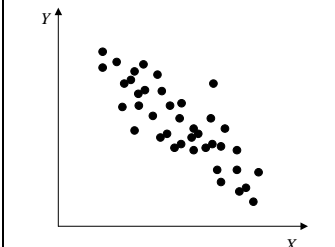
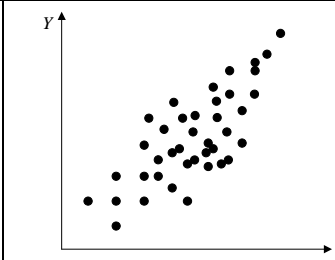
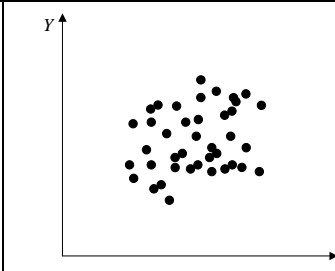
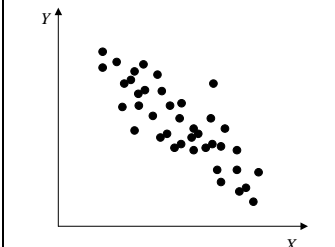
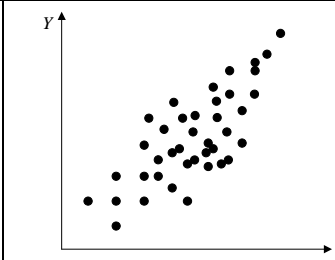
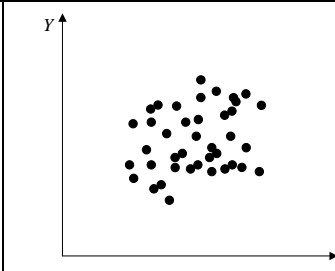
Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

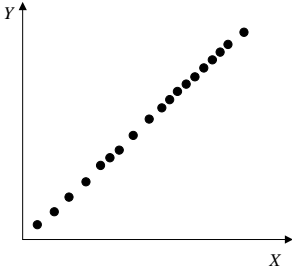
№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора

<p>1</p>	<p><b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b>  К какому типу шкалы относится номер студента в списке группы</p> <p>а) номинальной  б) измерений  в) порядковой  г) интервальной  д) относительной  е) дихотомической</p>	<p>УК-1.3.1</p>												
<p>2</p>	<p><b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b>  К какому типу шкалы относится город проживания сотрудника компании</p> <p>а) номинальной  б) измерений  в) порядковой  г) интервальной  д) относительной  е) дихотомической</p>	<p>УК-1.У.3</p>												
<p>3</p>	<p><b>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце. Запишите выбранные цифры под соответствующими буквами.</b>  Соотнесите точечную диаграмму данных с ее значением коэффициента корреляции <math>r</math></p> <table border="1" data-bbox="347 1108 1276 2011"> <thead> <tr> <th data-bbox="347 1108 817 1182">Точечная диаграмма данных</th> <th colspan="2" data-bbox="817 1108 1276 1182">Значение коэффициента корреляции <math>r</math></th> </tr> </thead> <tbody> <tr> <td data-bbox="347 1182 817 1444"> <p>1</p>  </td> <td data-bbox="817 1182 906 1444">А</td> <td data-bbox="906 1182 1276 1444"><math>r &gt; 0</math></td> </tr> <tr> <td data-bbox="347 1444 817 1720"> <p>2</p>  </td> <td data-bbox="817 1444 906 1720">Б</td> <td data-bbox="906 1444 1276 1720"><math>r &lt; 0</math></td> </tr> <tr> <td data-bbox="347 1720 817 2011"> <p>3</p>  </td> <td data-bbox="817 1720 906 2011">В</td> <td data-bbox="906 1720 1276 2011"><math>r = 1</math></td> </tr> </tbody> </table>	Точечная диаграмма данных	Значение коэффициента корреляции $r$		<p>1</p> 	А	$r > 0$	<p>2</p> 	Б	$r < 0$	<p>3</p> 	В	$r = 1$	<p>УК-1.У.3</p>
Точечная диаграмма данных	Значение коэффициента корреляции $r$													
<p>1</p> 	А	$r > 0$												
<p>2</p> 	Б	$r < 0$												
<p>3</p> 	В	$r = 1$												

4		Г	$r=0$																									
<p><b>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце. Запишите выбранные цифры под соответствующими буквами.</b></p> <p>Соотнесите величину коэффициента корреляции <math>r</math> с характеристикой силы связи между признаками</p>																												
<table border="1"> <thead> <tr> <th data-bbox="347 725 416 801">Величина коэффициента корреляции <math>r</math></th> <th data-bbox="416 725 810 801"></th> <th data-bbox="810 725 906 801">Характеристикой силы связи между признаками</th> <th data-bbox="906 725 1273 801"></th> </tr> </thead> <tbody> <tr> <td data-bbox="347 801 416 840">1</td> <td data-bbox="416 801 810 840">0.1-0.3</td> <td data-bbox="810 801 906 840">А</td> <td data-bbox="906 801 1273 840">высокая</td> </tr> <tr> <td data-bbox="347 840 416 878">2</td> <td data-bbox="416 840 810 878">0.3-0.5</td> <td data-bbox="810 840 906 878">Б</td> <td data-bbox="906 840 1273 878">слабая</td> </tr> <tr> <td data-bbox="347 878 416 916">3</td> <td data-bbox="416 878 810 916">0.5-0.7</td> <td data-bbox="810 878 906 916">В</td> <td data-bbox="906 878 1273 916">весьма высокая</td> </tr> <tr> <td data-bbox="347 916 416 954">4</td> <td data-bbox="416 916 810 954">0.7-0.9</td> <td data-bbox="810 916 906 954">Г</td> <td data-bbox="906 916 1273 954">умеренная</td> </tr> <tr> <td data-bbox="347 954 416 994">5</td> <td data-bbox="416 954 810 994">0.9-1.0</td> <td data-bbox="810 954 906 994">Д</td> <td data-bbox="906 954 1273 994">заметная</td> </tr> </tbody> </table>					Величина коэффициента корреляции $r$		Характеристикой силы связи между признаками		1	0.1-0.3	А	высокая	2	0.3-0.5	Б	слабая	3	0.5-0.7	В	весьма высокая	4	0.7-0.9	Г	умеренная	5	0.9-1.0	Д	заметная
Величина коэффициента корреляции $r$		Характеристикой силы связи между признаками																										
1	0.1-0.3	А	высокая																									
2	0.3-0.5	Б	слабая																									
3	0.5-0.7	В	весьма высокая																									
4	0.7-0.9	Г	умеренная																									
5	0.9-1.0	Д	заметная																									
5	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Найдите Моду в следующем наборе данных Стаж= {3,5,9,12,5}.</p> <p>Ответ запишите в числовом формате.</p> <p>Ответ:</p>			УК-1.В.1																								
6	<p><b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b></p> <p>Очистка данных – это:</p> <p>а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке данных.</p> <p>б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность анализа данных</p> <p>в) объект, содержащий структурированные данные, которые могут оказаться полезными для анализа данных</p> <p>г) комплекс методов и процедур, направленных на извлечение данных из различных источников и преобразования в единый формат для загрузки в хранилище данных или аналитическую систему.</p>			УК-2.3.3																								
7	<p><b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа .</b></p> <p>Обучающая выборка – это:</p> <p>а) группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов</p> <p>б) набор данных, каждая запись которого представляет собой учебный пример, содержащий входные признаки (данные), и соответствующий им правильный выходной результат</p> <p>в) набор данных, каждая запись которого представляет собой</p>																											

	учебный пример, содержащий заданное входное влияние г) выявление в сырых данных ранее неизвестных и практически полезных знаний, необходимых для принятия решений в различных сферах человеческой деятельности	
8	<b>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</b> Какие алгоритмы подходят для решения задачи, объекты в которой нужно разделить на классы: а) случайный лес; б) линейная регрессия; в) логистическая регрессия г) машина опорных векторов	УК-2.У.1
9	<b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b> Ошибка обучения – это: а) разность между желаемым и фактическим выходом модели на примерах обучающего множества б) ошибка, полученная на тестовом множестве в) имена, типы и метки исходной выборки данных г) набор данных, каждая запись которого представляет собой учебный пример	
10	<b>Прочитайте текст и запишите развернутый обоснованный ответ.</b> Найдите Интервал в следующем наборе данных Стаж= {3,5,9,12,5}. Ответ запишите в числовом формате. Ответ:	УК-2.В.3
11	<b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b> К какой стратегии обучения относится задача прогнозирования а) обучение с учителем б) обучение без учителя в) комбинированный метод г) другие	ПК-13.У.1
12	<b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b> Задача кластеризации это: а) задача машинного обучения, в которой метки объектов принимают ограниченное число значений б) задача машинного обучения, заключающаяся в объединении похожих объектов в однородные группы. в) задача машинного обучения, в которой метки объектов принимают любые численные значения г) задача машинного обучения, в которой оценивается отклонение ожидаемого и получаемого результата	
13	<b>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</b> Какая из этих задач является задачей классификации: а) прогноз температуры на завтра;	

	<p>б) разделение книг в библиотеке на научные и художественные;</p> <p>в) прогноз оценки студента по пятибалльной шкале на экзамене;</p> <p>г) определение центра и границ кластера.</p>	
14	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Назначение процедуры нормализации данных.</p> <p>Ответ:</p>	ПК-13.В.1
15	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Выполните постановку задачи классификации, используя следующие обозначения: <math>X = \{x_1, x_2, \dots, x_n\}</math> – множество примеров обучения, <math>Y = \{y_1, y_2\}</math> – множество меток класса</p> <p>Ответ:</p>	
16	<p><b>Прочитайте текст и выберите правильные ответы.</b></p> <p>Какие из следующих алгоритмов машинного обучения являются обучением с учителя:</p> <p>а) Random Forest;</p> <p>б) SVM;</p> <p>в) k-means;</p> <p>г) PCA.</p>	ПК-14.У.1
17	<p><b>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</b></p> <p>Выбрать число кластеров для алгоритма кластеризации «метод k-средних» можно:</p> <p>а) методом главных компонент;</p> <p>б) регуляризацией LASSO;</p> <p>в) силуэтным методом</p> <p>г) методом локтя</p>	
18	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Дайте определение понятию Обучающая выборка</p> <p>Ответ:</p>	
19	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Дайте определение понятию Ошибка обучения</p> <p>Ответ:</p>	
20	<p><b>Прочитайте текст и запишите развернутый обоснованный ответ.</b></p> <p>Дайте определение понятию Пример обучения</p> <p>Ответ:</p>	

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
1	Одномерный и двумерный анализ данных по известной выборке данных,

	публикованной на платформе Kaggle (соревновательная платформа для анализа данных). Выборка данных прилагается каждому варианту
2	Предобработка данных. опубликованных на платформе Kaggle. Выборка данных прилагается к каждому варианту.

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

## 11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала.

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

### Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

### Структура предоставления лекционного материала:

Раздел 1. Введение

Тема 1.1. Данные и их измерение.

Тема 1.2. Шкалы

Тема 1.3. Характеристика и классификация задач анализа данных

Тема 1.4. Этапы решения задачи анализа данных.

Раздел 2. Статистический анализ данных

Тема 2.1. Дескриптивный анализ

Тема 2.2. Анализ связей

Тема 2.3. Многомерный статистический анализ

Тема 2.4. Анализ временных рядов

Раздел 3. Кластерный анализ

Тема 3.1. Постановка задачи кластерного анализа



Тема 3.2. Метрики расстояний  
Тема Раздел 4. Классификация  
Тема 4.1. Постановка задачи классификации  
Тема 4.2. Методы классификации 3.3. Методы кластерного анализа  
Раздел 5. Заключение  
Тема 5.1. Перспективы развития анализа данных.

#### 11.2. Методические указания для обучающихся по прохождению практических занятий

Практическое занятие является одной из основных форм организации учебного процесса, заключающаяся в выполнении обучающимися под руководством преподавателя комплекса учебных заданий с целью усвоения научно-теоретических основ учебной дисциплины, приобретения умений и навыков, опыта творческой деятельности.

Целью практического занятия для обучающегося является привитие обучающимся умений и навыков практической деятельности по изучаемой дисциплине.

Планируемые результаты при освоении обучающимся практических занятий:

- закрепление, углубление, расширение и детализация знаний при решении конкретных задач;
- развитие познавательных способностей, самостоятельности мышления, творческой активности;
- овладение новыми методами и методиками изучения конкретной учебной дисциплины;
- выработка способности логического осмысления полученных знаний для выполнения заданий;
- обеспечение рационального сочетания коллективной и индивидуальной форм обучения.

#### Требования к проведению практических занятий

Результаты выполнения практических работ по требованию преподавателя выгружаются в личный кабинет или обсуждаются в аудитории во время занятий.

#### 11.3. Методические указания для обучающихся по выполнению лабораторных работ

В ходе выполнения лабораторных работ обучающийся должен углубить и закрепить знания, практические навыки, овладеть современной методикой и техникой эксперимента в соответствии с квалификационной характеристикой обучающегося. Выполнение лабораторных работ состоит из экспериментально-практической, расчетно-аналитической частей и контрольных мероприятий.

Выполнение лабораторных работ обучающимся является неотъемлемой частью изучения дисциплины, определяемой учебным планом, и относится к средствам, обеспечивающим решение следующих основных задач обучающегося:

- приобретение навыков исследования процессов, явлений и объектов, изучаемых в рамках данной дисциплины;
- закрепление, развитие и детализация теоретических знаний, полученных на лекциях;
- получение новой информации по изучаемой дисциплине;
- приобретение навыков самостоятельной работы с лабораторным оборудованием и приборами.

#### Задание и требования к проведению лабораторных работ

Задания к проведению лабораторных работ приведены в учебно-методическом пособии:

**Шифр 004 Т 23** Татарникова, Т. М. Анализ данных: учебно-методическое пособие / Т. М. Татарникова; С.-Петерб. гос. ун-т аэрокосм. приборостроения. - Санкт-Петербург: Изд-во ГУАП, 2021. - 121 с.

**Шифр 004 Б 76** Боженко, В.В. Язык программирования Python для анализа данных: учебно-методическое пособие / В. В. Боженко, Т. М. Татарникова; С.-Петерб. гос. ун-т аэрокосм. приборостроения. - Санкт-Петербург: Изд-во ГУАП, 2023. - 66 с.

#### Структура и форма отчета о лабораторной работе

1. Цель работы и содержание задания.
2. Описание хода выполнения работы.
3. Демонстрация полученных результатов.
4. Выводы о проделанной работе.
5. Ответы на контрольные вопросы.

#### Требования к оформлению отчета о лабораторной работе

Отчет оформляется согласно ГОСТ Р 2.105-2019. Обязательно наличие титульного листа и соответствие структуре отчета.

Отчет о проделанной работе выгружается в личный кабинет.

#### 11.4. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся, являются:

- учебно-методический материал по дисциплине;
- методические указания по выполнению контрольных работ (для обучающихся по заочной форме обучения).

#### 11.5. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Практические и лабораторные работы защищаются в течении двух недель после получения задания. Результаты защиты учитываются при проведении промежуточной аттестации.

#### 11.6. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя дифференцированный зачет.

Дифференцированный зачет – это форма оценки знаний, полученных обучающимся при изучении дисциплины с аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Студенты, успешно выполнившие все практические и лабораторные работы, и прошедшие тестирование на число баллов выше 85 баллов оцениваются на «отлично».

Студенты, успешно выполнившие все практические и лабораторные работы, и прошедшие тестирование на число баллов выше 69 баллов и не более 85 баллов оцениваются на «хорошо».

Студенты, успешно выполнившие все практические и лабораторные работы, и прошедшие тестирование на число баллов выше 54 баллов и не более 69 баллов оцениваются на «удовлетворительно».

Студенты, выполнившие не все практические и(или) лабораторные работы сдают дифф. зачет по вопросам, примерный перечень которых приведен в таблице 16.

Студенты, получившие оценку «удовлетворительно» по тестированию, имеют право сдать дифф. зачет и повысить оценку.

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой