

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)

«19» февраля 2025 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Машинное обучение и большие данные»
(Наименование дисциплины)

Код направления подготовки/ специальности	09.03.03
Наименование направления подготовки/ специальности	Прикладная информатика
Наименование направленности	Прикладная информатика в информационной сфере
Форма обучения	очная
Год приема	2025

Лист согласования рабочей программы дисциплины

Программу составил (а)

Проф., д-р техн. наук,
проф.

(должность, уч. степень, звание)



(подпись, дата)

Т.М. Татарникова

(инициалы, фамилия)

Программа одобрена на заседании кафедры № 41

«19» февраля 2025 г, протокол № 07-2024/25



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)

Аннотация

Дисциплина «Машинное обучение и большие данные» 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:

ПК-15 «Машинное обучение и большие данные»

Содержание дисциплины охватывает круг вопросов, связанных с изучением методов машинного обучения и их применения в задачах анализа больших данных.

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

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

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 в освоении компетенций по применению методов машинного обучения для извлечения знаний из больших 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
Профессиональные компетенции	ПК-15 Машинное обучение и большие данные	ПК-15.3.1 знать технологии и алгоритмы анализа данных ПК-15.У.1 уметь решать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных ПК-15.В.1 владеть языками программирования высокого уровня, ориентированных на работу с большими данными ПК-15.В.2 владеть навыками использования алгоритмов и моделей машинного обучения при решении прикладных задач

2. Место дисциплины в структуре ОП

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

- Введение в анализ данных
- Статистическая обработка информации
- Технологии программирования,
- Информационные системы и технологии.

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

- Организация научных исследований.

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
Общая трудоемкость дисциплины, 3Э/ (час)	3/ 108	3/ 108
Из них часов практической подготовки	34	34
Аудиторные занятия, всего час.	51	51

в том числе:		
лекции (Л), (час)	17	17
практические/семинарские занятия (ПЗ), (час)		
лабораторные работы (ЛР), (час)	34	34
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)	36	36
Самостоятельная работа , всего (час)	21	21
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Экз.	Экз.

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. Задачи машинного обучения	2		2		2
Тема 1.2. Интерпретация данных					
Раздел 2. Методы машинного обучения с учителем					
Тема 2.1. Классификация и регрессия	6		10		8
Тема 2.2. Алгоритмы машинного обучения					
Тема 2.3. Оценка неопределенности для классификаторов					
Раздел 3. Методы машинного обучения без учителя					
Тема 3.1. Предварительная обработка данных	4		10		6
Тема 3.2. Выделение признаков					
Тема 3.3. Кластеризация					
Раздел 4. Оценка качества модели					
Тема 4.1. Перекрестная проверка	4		8		3
Тема 4.2. Решетчатый поиск					
Тема 4.3. Метрики					
Раздел 5. Заключение					
Тема 5.1. Фреймворки и пакеты машинного обучения	1		4		2
Итого в семестре:	17		34		21
Итого	17	0	34	0	21

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	Тема 1.1. Задачи машинного обучения Контролируемое обучение. Процесс сбора данных. Неконтролируемое обучение. Понятие об обучающей выборке. Тема 1.2. Интерпретация данных Формат данных, признаки, измерение эффективности
Раздел 2	Тема 2.1. Классификация и регрессия Цель задачи классификация. Бинарная и мультиклассовая классификация. Сложность модели классификации. Тема 2.2. Алгоритмы машинного обучения Наивные байесовские классификаторы. Деревья решений. Ансамбли деревьев решений. Метод опорных векторов. Нейронные сети: прямого распространения, Хопфилда, рекуррентные, сверточные, Тема 2.3. Оценка неопределенности для классификаторов Решающая функция. Прогнозирование вероятностей.
Раздел 3	Тема 3.1. Предварительная обработка данных Виды предварительной обработки. Преобразование данных. Масштабирование обучающего и тестового набора данных. Тема 3.2. Выделение признаков Анализ главных компонент. Факторизация. Тема 3.3. Кластеризация Нейронная сеть Кохонена.
Раздел 4	Тема 4.1. Перекрестная проверка Кросс-проверка. Стратегии перекрестной проверки. Тема 4.2. Решетчатый поиск Простой решетчатый поиск. Решетчатый поиск с перекрестной проверкой. Экономичный решетчатый поиск Тема 4.3. Метрики Типы ошибок. Матрица ошибок. Точность, полнота, F-мера. ROC-кривые
Раздел 5	Тема 5.1. Фреймворки и пакеты машинного обучения Знакомство с библиотеками Python: NumPy, SciPy, Pandas, Matplotlib scikit-learn, statsmodel

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
Учебным планом не предусмотрено					

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	Наивные байесовские	4	2	2,4,5

	классификаторы.			
2	Классификация методом дерева решений.	4	1	2,4,5
3	Классификация методом случайного леса.	4	1	
4	Интерполяция с применением нейронной сети прямого распространения	4	1	2,4,5
5	Распознавание образов с применением нейронной сети Хопфилда	4	1	2,4,5
6	Кластеризация с применением нейронной сети Кохонена	4	1	3,4,5
7	Анализ текста с применением рекуррентных нейронных сетей	4	1	3,4,5
8	Классификация с применением сверточной нейронной сети	4	1	3,4,5
	Итоги текущего контроля.	2	1	1-5
Всего		34	10	

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)	16	30
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)	4	20
Домашнее задание (ДЗ)		
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)	3	7
Всего:	21	21

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
--------------	--------------------------	--

адрес		электронных 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.-Петерб. 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. http://www.machinelearning.ru 2. https://www.coursera.org/learn/machine-learning 3. https://www.coursera.org/specializations/machine-learning-data-analysis https://it.mail.ru/video/playlists/16/	1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. 2. Machine learning by Andrew Ng, 3. Курс «Анализ данных» от школы анализа данных, 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
Экзамен	Список вопросов к экзамену; Экзаменационные билеты; Задачи; Тесты.

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»	– обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	– обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения;

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
	– владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	– обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	– обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений.

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 зачета / дифф. зачета	Код индикатора
1	Методы машинного обучения	УК-15.3.1
2	Переобучение, недообучение, предобучение	
3	Ошибка обучения	
4	Линейные модели классификации	УК-15.У.1
5	Обучение линейных классификаторов	
6	Цель задачи классификация.	
7	Бинарная классификация	
8	Мультиклассовая классификация.	
9	Сложность модели классификации.	
10	Наивные байесовские классификаторы.	ПК-15.В.1
11	Деревья решений.	
12	Ансамбли деревьев решений.	
13	Метод опорных векторов	
14	Методы классификации.	УК-15.3.1
15	Оценка ошибка классификации	УК-15.У.1
16	Решающая функция.	
17	Прогнозирование вероятностей	ПК-15.В.1
18	Нейронные сети прямого распространения	ПК-15.3.1
19	Нейронная сеть Хопфилда	
20	Сверточные нейронные сети	
21	Рекуррентные нейронные сети	
22	Нейронная сеть Кохонена	
23	Предобработка данных.	ПК-15.В.1
24	Постановка задачи классификации	
25	Этапы решения задач обучения без учителя	ПК-15.3.1
26	Кросс-проверка.	ПК-15.У.1
27	Стратегии перекрестной проверки.	

28	Кросс-проверка.	
29	Стратегии перекрестной проверки.	
30	Простой решетчатый поиск.	ПК-15.В.2
31	Решетчатый поиск с перекрестной проверкой.	
32	Экономичный решетчатый поиск	

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
	Учебным планом не предусмотрено	

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

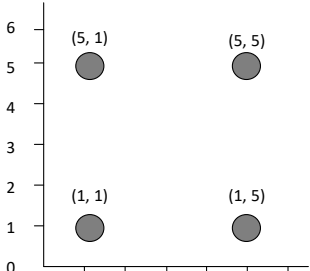
Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

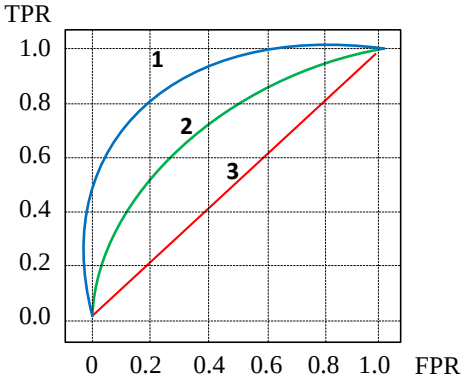
№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

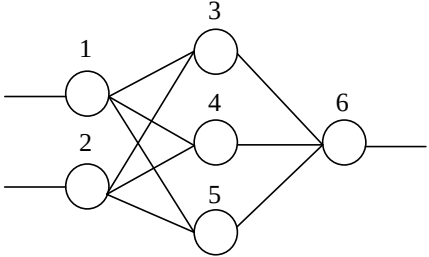
Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора
1	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Определения дохода клиента банка является задачей: а) регрессии; б) классификации; в) кластеризации г) сегментации.	УК-15.3.1
2	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Какая функция в логистической регрессии переводит значения в интервал $[0, 1]$: а) $\tan(x)$; б) $1/(1 + e^{-x})$; в) $\log(x)$; г) $\sin(x)$.	
3	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов 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	УК-15.У.1

4	г) сегментация объектов по схожим признакам Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Какой из следующих методов можно использовать для сокращения размерности: а) опорных векторов; б) главных компонент; в) ближайшего соседа; г) k-ближайших соседей.	
5	Прочитайте текст и запишите развернутый обоснованный ответ Определите координаты центроида кластера. Результат запишите в числовом формате в виде (x,y) 	УК-15.В.2
6	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Какая из этих задач является задачей классификации: а) прогноз температуры на завтра; б) разделение книг в библиотеке на научные и художественные; в) прогноз оценки студента по пятибалльной шкале на экзамене; г) выявление признаков, влияющих на успеваемость студента.	УК-15.3.1
7	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Сколько коэффициентов нужно для построения модели линейной регрессии для одного признака: а) 1; б) 3; в) 2; г) 0.	
8	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов Борьба с переобучением предполагает следующие механизмы: а) увеличить данные в наборе. б) уменьшить число параметров модели. в) регуляризация г) усложнение модель за счет добавления параметров д) уменьшение регуляризации	УК-15.У.1
9	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов Борьба с недообучением предполагает следующие механизмы: а) увеличить данные в наборе.	

	б) уменьшить число параметров модели. в) регуляризация г) усложнение модель за счет добавления параметров д) уменьшение регуляризации	
10	Прочитайте текст и запишите развернутый обоснованный ответ На графике ниже приведены ROC-кривые, оценивающие эффективность трех классификаторов 1, 2, 3. Какой из классификаторов 1 или 2 или 3 точнее?  The graph shows True Positive Rate (TPR) on the y-axis and False Positive Rate (FPR) on the x-axis, both ranging from 0.0 to 1.0. Three curves are plotted: a blue curve labeled '1' which is the highest, a green curve labeled '2' which is below the blue one, and a red diagonal line labeled '3' representing a random classifier.	УК-15.В.1
11	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Что может повысить качество классификации в Random Forest: а) использование кластеров; б) использование большого количества ансамблей деревьев; в) использование ковариационной матрицы; г) применение метода PCA.	ПК-15.3.1
12	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Пусть есть задача с 2 классами: $Y = \{-1, +1\}$. К какому классу будет относиться результат, если $f(X, W) > 0$ а) +1; б) -1; в) обоим классам; г) ни к какому из перечисленных.	
13	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов Какие методы машинного обучения относятся к стратегии обучения с учителем: а) классификация; б) регрессия; в) кластеризация; г) ассоциация.	
14	Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо Расположите следующие этапы построения классификатора методом машинного обучения:	

	А – обучение классификатора; Б – тестирование классификатора В – использование модели (классификация новых объектов); Г – формализация модели классификатора (выбор метода машинного обучения)	
15	Прочитайте текст и запишите развернутый обоснованный ответ На рисунке ниже приведена архитектура искусственной нейронной сети (ИНС) прямого распространения со следующими исходными данными:  Рис. Архитектура ИНС прямого распространения - матрица весовых коэффициентов: $\begin{pmatrix} 1 & 0 & -1 & 0,5 & -0,5 & 0 \\ 0 & 0,2 & -0,7 & 0,3 & 0,2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,1 \\ 0 & 0 & 0 & 0 & 0 & 0,1 \\ 0 & 0 & 0 & 0 & 0 & -0,3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ - Функция активации сигмоидальная Определить выходное значение 1-го нейрона и записать в числовом формате с десятичной точкой с двумя знаками после запятой	ПК-15.В.1
16	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов Как называется задача машинного обучения, направленная на предсказание значения той или иной числовой величины для входных данных: а) обучение без учителя; б) обучение с учителем; в) регрессии; г) прогнозирования д) кластеризации.	ПК-15.3.1
17	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа Что является причиной недообучения: а) использовании слишком простых моделей; б) использовании избыточно сложных моделей; в) кросс-валидация; г) снижение значения ошибки обучения.	
18	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов	ПК-15.У.1

19	Какие из следующих алгоритмов машинного обучения являются обучением без учителя: а) Random Forest; б) SVM; в) k-means; г) нейронная сеть Кохонена д) все выше перечисленные. Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов Какие методы машинного обучения относятся к стратегии обучения без учителем: а) классификация; б) регрессия; в) кластеризация; г) ассоциация.																					
20	Прочитайте текст и запишите развернутый обоснованный ответ Дан временной ряд (число заболевших гриппом, человек/сутки): <table><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>32</td><td>103</td><td>589</td><td>1201</td><td>3574</td><td>2800</td><td>2521</td><td>2005</td><td>1693</td><td>1278</td></tr></table> Спрогнозировать методом kNN при k=3 число заболевших на 1 шаг вперед. Результат запишите в числовом формате с десятичной точкой (два знака после запятой).	1	2	3	4	5	6	7	8	9	10	32	103	589	1201	3574	2800	2521	2005	1693	1278	ПК-15.B.2
1	2	3	4	5	6	7	8	9	10													
32	103	589	1201	3574	2800	2521	2005	1693	1278													

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

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

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:

Тема 1.1. Задачи машинного обучения

Контролируемое обучение. Процесс сбора данных. Неконтролируемое обучение.

Понятие об обучающей выборке.

Тема 1.2. Интерпретация данных

Формат данных, признаки, измерение эффективности

Тема 2.1. Классификация и регрессия

Цель задачи классификация. Бинарная и мультиклассовая классификация.

Сложность модели классификации.

Тема 2.2. Алгоритмы машинного обучения

Наивные байесовские классификаторы. Деревья решений. Ансамбли деревьев решений. Метод опорных векторов. Нейронные сети: прямого распространения, Хопфилда, рекуррентные, сверточные,

Тема 2.3. Оценка неопределенности для классификаторов

Решающая функция. Прогнозирование вероятностей.

Тема 3.1. Предварительная обработка данных

Виды предварительной обработки. Преобразование данных. Масштабирование обучающего и тестового набора данных.

Тема 3.2. Выделение признаков

Анализ главных компонент. Факторизация.

Тема 3.3. Кластеризация

Нейронная сеть Кохонена.

Тема 4.1. Перекрестная проверка

Кросс-проверка. Стратегии перекрестной проверки.

Тема 4.2. Решетчатый поиск

Простой решетчатый поиск. Решетчатый поиск с перекрестной проверкой.

Экономичный решетчатый поиск

Тема 4.3. Метрики

Типы ошибок. Матрица ошибок. Точность, полнота, F-мера. ROC-кривые

Тема 5.1. Фреймворки и пакеты машинного обучения

Знакомство с библиотеками Python: NumPy, SciPy, Pandas, Matplotlib scikit-learn, statsmodel.

11.2. 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. Методы машинного обучения: учебное пособие / Т. М. Татарникова, В. В. Боженко; С.-Петерб. гос. ун-т аэрокосм. приборостроения. - Санкт-Петербург: Изд-во ГУАП, 2023. - 100 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.

11.3. 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).

11.4. 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.

Текущий контроль осуществляется через защиту результатов лабораторных работ.

Лабораторные работы защищаются в течение двух недель после получения задания. Результаты защиты учитываются при проведении промежуточной аттестации.

11.5. Методические указания для обучающихся по прохождению промежуточной аттестации.

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

- экзамен – форма оценки знаний, полученных обучающимся в процессе изучения всей дисциплины или ее части, навыков самостоятельной работы, способности применять их для решения практических задач. Экзамен, как правило, проводится в период экзаменационной сессии и завершается аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

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

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

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

Студенты, выполнившие не все лабораторные работы сдают экзамен по вопросам, примерный перечень которых приведен в таблице 16.

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