

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

Кафедра № 32

УТВЕРЖДАЮ

Руководитель образовательной программы

ДОЦ., К.Т.Н., ДОЦ.

(должность, уч. степень, звание)

О.Я. Солёная

(инициалы, фамилия)


(подпись)

«27» июня 2024 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Машинное обучение и анализ данных»
(Наименование дисциплины)

Код направления подготовки	13.04.02
Наименование направления подготовки	Электроэнергетика и электротехника
Наименование направленности	Цифровая энергетика
Форма обучения	очная
Год приема	2024

Лист согласования рабочей программы дисциплины

Программу составил

проф., д.т.н., доц.
(должность, уч. степень, звание)


(подпись, дата)

С.А. Сериков
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 32

«26» июня 2024 г., протокол № 10

Заведующий кафедрой № 32

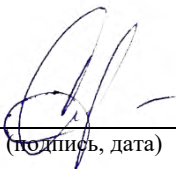
к.т.н., доц.
(уч. степень, звание)


(подпись, дата)

С.В. Солёный
(инициалы, фамилия)

Заместитель директора института №3 по методической работе

ст. преподаватель
(должность, уч. степень, звание)


(подпись, дата)

Н.В. Решетникова
(инициалы, фамилия)

Аннотация

Дисциплина «Машинное обучение и анализ данных» входит в образовательную программу высшего образования – программу магистратуры по направлению подготовки 13.04.02 «Электроэнергетика и электротехника» направленности «Цифровая энергетика». Дисциплина реализуется кафедрой «№32».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-1 «Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий»

ОПК-2 «Способен применять современные методы исследования, оценивать и представлять результаты выполненной работы»

ПК-3 «способен применять технологии цифровых двойников для информационного моделирования объектов профессиональной деятельности.»

Содержание дисциплины охватывает круг вопросов, связанных с применением методов машинного обучения и интеллектуального анализа данных при решении прикладных задач в области электроэнергетики и электротехники.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, 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. Цель преподавания дисциплины: формирование у студентов комплекса профессиональных знаний и умений в области машинного обучения и интеллектуального анализа данных.

1.2. Дисциплина входит в состав обязательной части 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 критический анализ ситуаций на основе системного подхода, выработать стратегию действий	УК-1.3.2 знать цифровые ресурсы, инструменты и сервисы, включая интеллектуальные, для решения задач/проблем профессиональной деятельности УК-1.У.1 уметь искать нужные источники информации; анализировать, сохранять и передавать информацию с использованием цифровых средств; выработать стратегию действий для решения проблемной ситуации УК-1.В.2 владеть навыками использования алгоритмов и цифровых средств, предназначенных для анализа информации и данных
Общепрофессиональные компетенции	ОПК-2 Способен применять современные методы исследования, оценивать и представлять результаты выполненной работы	ОПК-2.Д.2 выбирает и применяет современные средства представления и обработки данных, использует алгоритмы машинного обучения ОПК-2.Д.4 знает методы разработки алгоритмов и программного обеспечения в рамках систем искусственного интеллекта
Профессиональные компетенции	ПК-3 способен применять технологии цифровых двойников для информационного моделирования объектов профессиональной деятельности.	ПК-3.Д.1 формирует классифицированную базу данных о состоянии электроэнергетических объектов и систем для построения информационных моделей ПК-3.Д.4 использует методы разработки оригинальных алгоритмов и программных решений с использованием современных технологий

2. Место дисциплины в структуре ОП

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

- «Математические методы и модели в научных исследованиях»,
- «Электрические системы и сети».

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

- «Интеллектуальные системы энергосбережения»,
- «Цифровые двойники в электроэнергетике»,
- «Электрические системы и сети».

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
		№1
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	3/ 108	3/ 108
Из них часов практической подготовки	5	5
Аудиторные занятия, всего час.	34	34
в том числе:		
лекции (Л), (час)	17	17
практические/семинарские занятия (ПЗ), (час)	17	17
лабораторные работы (ЛР), (час)		
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
Самостоятельная работа, всего (час)	74	74
Вид промежуточной аттестации: 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 1					
Раздел 1. Основные понятия машинного обучения	2	2			12
Раздел 2. Основные модели машинного обучения.	4	4			15
Раздел 3. Реализация алгоритмов машинного обучения.	4	4			16

Раздел 4. Предобработка данных в задачах машинного обучения.	4	4			16
Раздел 5. Оптимизация потоков операций при помощи конвейеров.	3	3			15
Итого в семестре:	17	17			74
Итого	17	17	0	0	74

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
1	<p>Основные понятия машинного обучения Предмет дисциплины и ее задачи. Структура и содержание дисциплины. Взаимосвязь изучаемой дисциплины с другими дисциплинами учебного плана. Характеристика основной и дополнительной литературы</p> <p>Общая характеристика методов и разделов, входящих в предметную область – машинное обучение. Область применения и задачи, решаемые технологиями машинного обучения, основные понятия и обозначения. Основные библиотеки языка Python для работы с данными.</p> <p>Три вида машинного обучения:</p> <ul style="list-style-type: none"> – обучение с учителем. Основные решаемые задачи. Области применения. <p>Прогнозирование меток классов в задачах классификации. Прогнозирование значений непрерывной целевой переменной на основе регрессионного анализа;</p> <ul style="list-style-type: none"> – обучение с подкреплением. Основные решаемые задачи. Области применения. <p>Основные модели обучения с подкреплением;</p> <ul style="list-style-type: none"> – обучение без учителя. Основные решаемые задачи. Области применения. Работа с немаркированными данными – кластерный анализ. Обнаружение скрытых структур в данных и иерархии кластеров. Снижение размерности для «сжатия данных». <p>Выявление ассоциативных связей.</p>
2	<p>Основные модели машинного обучения.</p> <p>Основные этапы машинного обучения, их назначение и основные особенности:</p> <ul style="list-style-type: none"> – предварительная обработка данных в системах машинного обучения. <p>Назначение. Основные особенности;</p> <ul style="list-style-type: none"> – этап «Обучение» в системах машинного обучения. Принципы формирования тренировочного набора данных. Использование перекрёстной проверки. <p>Гиперпараметрическая оптимизация. Основные метрики качества машинного обучения;</p> <ul style="list-style-type: none"> – этап оценки обобщающей способности в системах машинного обучения. <p>Понятия точности (precision) и верности (accuracy) функционирования моделей.</p> <p>Особенности выбора оптимальной модели машинного обучения. Типовая зависимость между сложностью модели машинного обучения и оценкой её эффективности. Типовая зависимость оценки эффективности моделей машинного обучения от размера обучающей последовательности.</p> <p>Формальная постановка задачи машинного обучения</p>

3	<p>Реализация алгоритмов машинного обучения. Модули NumPy, Pandas, Matplotlib, Scikit-Learn. Строение и функции биологического нейрона. Реализация перцептрона на Python с использованием scikit-learn. Тренировка перцептрона. Адаптивный линейный нейрон (ADALINE). Тренировка ADALINE методами пакетного и стохастического градиентного спуска. Динамическое обучение. Модель логистической регрессии. Тренировка логистической регрессионной модели в scikit-learn. Решение проблемы переобучения при помощи регуляризации. Метод опорных векторов (SVM). Классификация с максимальным зазором и с мягким зазором на основе SVM. Решение нелинейных задач ядерным методом SVM. Модели машинного обучения на основе деревьев решений. Объяснение результатов обучения. Случайный лес. Модель машинного обучения на основе k ближайших соседей. Наивный Байесовский классификатор. Композиции алгоритмов машинного обучения (ансамблевое обучение). Комитет большинства, бэггинг, бустинг, стеккинг.</p>
4	<p>Предобработка данных в задачах машинного обучения. Решение проблемы пропущенных данных. Обработка категориальных данных. Разбивка набора данных на тренировочное и тестовое подмножества. Приведение признаков к одинаковой шкале. Отбор содержательных признаков. Определение важности признаков. Сжатие данных путем снижения размерности. Снижение размерности без учителя на основе анализа главных компонент. Сжатие данных с учителем путем линейного дискриминантного анализа. Ядерный метод анализа главных компонент для нелинейных отображений.</p>
5	<p>Оптимизация потоков операций при помощи конвейеров. Тонкая настройка машинно-обучаемых моделей методом сеточного поиска. Тренировка искусственных нейронных сетей для распознавания изображений. Основные тенденции развития технологий машинного обучения</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 1					
1	Разработка модели линейного адаптивного нейрона (ADALINE) на основе базового класса перцептрона	групповая	2	2	1
2	Классификация на основе перцептронной модели с использованием библиотеки Scikit-learn	групповая	3	3	2, 3
3	Классификация на основе модели логистической	групповая	3	3	2-5

	регрессии с использованием библиотеки Scikit-learn				
4	Классификация методом опорных векторов с использованием библиотеки Scikit-learn	групповая	3	3	2-5
5	Классификация линейно разделимых объектов (три признака, четыре класса) с использованием библиотеки Scikit-learn	групповая	3	3	2-5
6	Классификация линейно неразделимых объектов с использованием библиотеки Scikit-learn	групповая	3	3	2-5
Всего			17	17	

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
Учебным планом не предусмотрено				
Всего				

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 1, 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)		
Подготовка к промежуточной аттестации (ПА)	14	14
Всего:	74	74

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 адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
УДК 681.3.07	Рашка Себастьян, Мирджалили Вахид. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2, 3-е изд. Пер. с англ. – СПб.: «Диалектика», 2020. – 848 с.	
УДК: 681.3.07	Жерон Орельен. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем, 2-е изд.: Пер. с англ.– СПб.: ООО «Диалектика», 2020.–1040 с.	
УДК 004.43 ПЗ7	Плас Дж. Вандер. Python для сложных задач: наука о данных. 2-е междунар. изд. – Астана: «Спринт Бук», 2024. – 592 с.	
УДК 681.3.07	Джоши Прадик. Искусственный интеллект с примерами на Python.: Пер. с англ. - СПб.: ООО "Диалектика", 2019. – 448 с.	
ISBN 978-5- 9908910-8-1	Андреас Мюллер, Сара Гвидо. Введение в машинное обучение с помощью python. Руководство для специалистов по работе с данными. – М.: ИЦ «Гевисста», 2017. – 393 с.	
УДК 0004.438	Луис Педро Коэльо, Вилли Ричарт. Построение систем машинного обучения на языке Python. 2-е издание/ пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 302 с.	
УДК 681.3.07	Элбон Крис. Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. — СПб.: БХВ-Петербург, 2019. — 384 с.: ил.	
УДК 004.62	Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и науки о данных. – СПб.: Питер, 2017. – 336 с.	

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
http://www.machinelearning.ru/wiki/index.php?title=Заглавная_страница	Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных
http://python.org	Официальный сайт языка Python
https://www.anaconda.com/download/	Anaconda - Дистрибутив Python,

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	Anaconda 2020.11 For Windows Installer
2	Python 3.7 version
3	Spyder 4.2.3

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
1	Онлайн-книга «Справочник по науке о данных Python» https://jakevdp.github.io/PythonDataScienceHandbook/05.06-linear-regression.html
2	Датасеты по машинному обучению http://archive.ics.uci.edu/ml/datasets.php

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	Б.М. 21-21
2	Компьютерный класс	Б.М. 31-04

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 зачет	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 5-балльная шкала	Характеристика сформированных компетенций
«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; – 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.

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – 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 экзамена	Код индикатора
	Учебным планом не предусмотрено	

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

Таблица 16 – Вопросы (задачи) для дифф. зачета

№ п/п	Перечень вопросов (задач) для дифф. зачета	Код индикатора
	<ol style="list-style-type: none"> 1. Дайте определение понятию «Искусственный интеллект». Поясните термины «Слабый искусственный интеллект» и «Сильный искусственный интеллект». Перечислите особенности слабоформализованных задач как основной области применения технологий искусственного интеллекта. Приведите примеры таких задач. 2. По какому признаку можно выделить интеллектуальные задачи среди других задач обработки информации? Приведите примеры интеллектуальных задач. Поясните, что называют интеллектуальными информационными технологиями. 3. Дайте определение понятию «Машинное обучение». С чем связано свойство обучения в таких моделях? Перечислите основные факторы, влияющие на эффективность использования методов машинного обучения, и дайте им краткую характеристику. 4. Приведите классификацию систем машинного обучения по объёму и типу контроля во время обучения, в зависимости от способности обучаться на основе потока входящих данных, по способу обобщения знаний. Дайте краткую характеристику перечисленным классам систем. 5. Поясните особенности обучения с учителем как вида машинного обучения. Приведите постановку задачи обучения с учителем. Перечислите типы задач обучения с учителем, в зависимости от вида множества допустимых ответов. Поясните, что называется признаковым описанием объекта из множества исследуемых объектов. Какие типы признаков существуют? 	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
	<ol style="list-style-type: none"> 6. Перечислите основные этапы решения задачи машинного обучения с учителем. Поясните содержание каждого этапа. 7. Поясните особенности обучения без учителя как вида машинного обучения. Перечислите наиболее характерные области 	<p>ОПК-2.Д.2 ОПК-2.Д.4</p>

	<p>использования обучения без учителя. Какие типы задач решаются методами обучения без учителя? Дайте краткую характеристику этим задачам.</p> <p>8. Поясните особенности обучения с подкреплением как вида машинного обучения. Приведите постановку задачи обучения с подкреплением. Опишите схему взаимодействия агента со средой при реализации обучения с подкреплением. Перечислите характерные области применения обучения с подкреплением.</p> <p>9. Поясните особенности оценки качества работы классификатора. Почему коэффициент ошибочных классификаций и правильность классификации не всегда являются эффективными показателями качества классификации? Что такое матрица неточностей? Как определить точность и полноту классификации с использованием матрицы неточностей? Какие особенности классификатора характеризуют точность и полнота?</p> <p>10. Как выглядят типовые зависимости точности от полноты классификации, а также точности и полноты от порога принятия решения? Поясните такое поведение данных зависимостей. Поясните, что такое «мера F_1», как она определяется и для чего используется.</p> <p>11. Поясните, как используется кривая рабочей характеристики приёмника (ROC) для оценки качества работы классификатора. Приведите характерный вид данной кривой. Как выглядит кривая рабочей характеристики приёмника при чисто случайной классификации? Для каких целей обычно используется измерение площади под кривой рабочей характеристики приёмника (AUC)?</p> <p>12. Поясните, как строится матрица неточностей при многоклассовой классификации. Как можно оценить точность и полноту для каждого класса, а также результирующие значения точности и полноты классификации.</p> <p>13. Поясните особенности оценки качества работы моделей регрессии. В каких случаях отдают предпочтение критерию качества RMSE, а в каких случаях критерию MAE?</p> <p>14. В чём состоит свойство обобщающей способности модели машинного обучения. Какие приёмы используются для теоретической и эмпирической оценки обобщающей способности: оценка математического ожидания средних потерь, оценка вероятности переобучения, оценка на отложенных данных (<i>hold-out</i>), оценка скользящего контроля (<i>cross-validation</i>).</p> <p>15. Назовите основные причины возникновения переобучения и недообучения в моделях машинного обучения. Перечислите основные способы минимизации переобучения.</p> <p>16. Поясните понятия смещения, дисперсии и неустранимой погрешности как трёх составляющих ошибки обобщения моделей машинного обучения.</p> <p>17. Поясните характер изменения оценок эффективности моделей машинного обучения при возрастании их сложности (<i>кривые проверки, validation curve</i>) и размера обучающей выборки (<i>кривые обучения, learning curve</i>).</p>	
	<p>18. Персептрон Розенблатта. Структурная схема и правило обучения персептрона.</p>	<p>ПК-3.Д.1 ПК-3.Д.4</p>

<p>19. Модель адаптивного линейного нейрона. Обучение ADALINE при помощи метода градиентного спуска. Пакетный и стохастический градиентный спуск.</p> <p>20. Модель логистической регрессии. Покажите, что выход сигмоидальной функции можно интерпретировать как вероятность принадлежности отдельно взятого образца первому классу.</p> <p>21. Обучение модели логистической регрессии. Решение проблемы переобучения при помощи регуляризации. L2-регуляризация</p> <p>22. Метод опорных векторов. Классификация с мягким зазором для обеспечения сходимости оптимизации при допустимости ошибочной классификации.</p> <p>23. Решение нелинейных задач при помощи метода опорных векторов. Добавление к набору данных дополнительных признаков. Использование признаков близости. RBF-функция как функция подобия.</p> <p>24. Гиперпараметрическая оптимизация моделей машинного обучения. Использование решётчатого поиска для выбора наилучших гиперпараметров модели.</p> <p>25. Модель на основе дерева принятия решений. Основные особенности модели. Построение дерева решений на основе принципа наибольшего прироста информации. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>26. Модель на основе случайного леса. Особенности модели. Алгоритм «выращивания» случайного леса. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>27. Классификатор на основе k ближайших соседей и его особенности. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>28. Наивный байесовский классификатор. Основные особенности модели и области её применения.</p> <p>29. Приёмы предварительной обработки данных в моделях машинного обучения. Решение проблемы с недостающими данными, обработка категориальных (<i>именных и порядковых</i>) данных, выбор значимых признаков (<i>использование L1 регуляризации, алгоритм последовательного выбора признаков, оценка важности признаков с помощью случайных лесов</i>).</p> <p>30. Выделение признаков: сжатие данных с помощью понижения размерности пространства признаков. Анализ главных компонент (<i>Principal Component Analysis</i>). Линейный дискриминантный анализ (<i>Linear Discriminant Analysis</i>).</p> <p>31. Ансамбли моделей машинного обучения. Основные подходы к созданию ансамблей. Бэггинг (<i>bagging</i>).</p> <p>32. Ансамблевое обучение. Бустинг (<i>boosting</i>). Алгоритмы Boost1, адаптивный бустинг (<i>AdaBoost</i>), градиентный бустинг (<i>GBM</i>).</p> <p>33. Ансамблевое обучение. Стекинг (<i>stacking</i>).</p>	
--	--

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 тип. Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора</p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p>		
1	<p>Задача классификации - это ...</p> <p>А) поиск группы, в которую можно объединить объекты, имеющих схожие характеристики</p> <p>Б) определение зависимости между двумя числовыми переменными</p> <p>В) предсказание числового значения целевой переменной на основе значений независимых переменных</p> <p>Г) поиск класса, к которому относится объект, на основе его признаков</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
2	<p>Задача регрессии - это ...</p> <p>А) предсказание числового значения целевой переменной на основе значений независимых переменных</p> <p>Б) поиск класса, к которому относится объект, на основе его признаков</p> <p>В) описание связи между двумя категориальными переменными</p> <p>Г) сравнительный анализ нескольких числовых переменных с помощью графиков</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
3	<p>В каком из случаев имеет смысл сразу исключить колонку данных:</p> <p>А) Вся колонка имеет одно и то же значение;</p> <p>Б) В колонке каждое значение - уникальное (вещественные числа);</p> <p>В) Колонка имеет два уникальных значения;</p> <p>Г) Колонка содержит строки.</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
4	<p>Какое из направлений исследований искусственного интеллекта использует случайный подбор, комбинирование и вариации искомых параметров с применением механизмов, напоминающих биологическую эволюцию?</p> <p>А) Искусственные нейронные сети;</p> <p>Б) Классическое машинное обучение;</p> <p>В) Генетические алгоритмы;</p>	<p>ОПК-2.Д.2 ОПК-2.Д.4</p>

	Г) Системы нечёткого вывода	
5	<p>Гиперпараметры модели машинного обучения – это:</p> <p>А) параметры, которые определяют зависимости между входными данными и выходами</p> <p>Б) параметры, которые определяют критерии остановки обучения модели</p> <p>В) параметры, которые определяют количество нейронов в скрытых слоях нейронной сети</p> <p>Г) параметры, которые не связаны непосредственно с обучением модели, но влияют на ее качество</p>	ОПК-2.Д.2 ОПК-2.Д.4
6	<p>Метрика, которая не подходит для оценки качества регрессионных моделей:</p> <p>А) R-квадрат (коэффициент детерминации)</p> <p>Б) MSE</p> <p>В) F1-мера</p> <p>Г) MAE</p>	ПК-3.Д.1 ПК-3.Д.4
7	<p>Преимущество библиотеки <code>numpy</code> перед использованием обычных списков в Python при работе с большими массивами данных:</p> <p>А) Легко добавлять/удалять элементы из списка</p> <p>Б) Позволяет быстро выполнять математические операции над массивами</p> <p>В) Обеспечивает простой доступ к элементам массива</p> <p>Г) Выполняет конвертацию между различными форматами данных.</p>	ПК-3.Д.1 ПК-3.Д.4
<p>2 тип. Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора</p> <p>Инструкция: Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов</p>		
8	<p>Задачи, из нижеперечисленных, являются задачами прогнозирования:</p> <p>А) математический прогноз даты сильных землетрясений</p> <p>Б) прогнозирование вероятности летального исхода</p> <p>В) определение длительности и исхода заболевания</p> <p>Г) обнаружение спама</p> <p>Д) задачи поискового вывода</p>	УК-1.3.2 УК-1.У.1 УК-1.В.2
9	<p>Что такое машинное обучение?</p> <p>А) Это процесс, в ходе которого система обрабатывает большое число примеров, выявляет закономерности и использует их, чтобы прогнозировать характеристики новых данных;</p>	УК-1.3.2 УК-1.У.1 УК-1.В.2

	<p>Б) Особенность такого обучения заключается в том, что данные представлены с известными правильными ответами, т.е. данные с метками;</p> <p>В) Область математики, изучающая алгоритмы для нахождения экстремумов функций;</p> <p>Г) Область искусственного интеллекта, связанная с обучением компьютеров на данных;</p> <p>Д) Область практической деятельности и научных исследований самообучающихся алгоритмов, извлекающих смысл (знания) из данных, средство создания моделей алгоритмов.</p> <p>Е) Методика обработки изображений.</p>	
10	<p>Как бороться с эффектом переобучения?</p> <p>А) Сделать модель проще (менее комплексной);</p> <p>Б) Взять меньше обучающих данных;</p> <p>В) Накопить больше обучающих данных;</p> <p>Г) Сделать модель сложнее (более комплексной);</p> <p>Д) Повысить качество обучающей выборки путём снижения шума в данных, исправления ошибок, устранения выбросов и т.д.;</p> <p>Е) Наложить ограничения на пространство параметров модели;</p> <p>Ж) Проверять метрики только по обучающей выборке.</p>	<p>УК-1.3.2</p> <p>УК-1.У.1</p> <p>УК-1.В.2</p>
11	<p>Возможными решениями проблемы недообучения являются:</p> <p>А) Выбор более сложной модели с большим числом настраиваемых параметров;</p> <p>Б) Предоставление обучающему алгоритму лучших признаков (конструирование признаков);</p> <p>В) Уменьшение ограничений модели (например, снижение величины гиперпараметра регуляризации).</p> <p>Г) Ни один из перечисленных.</p>	<p>ОПК-2.Д.2</p> <p>ОПК-2.Д.4</p>
12	<p>Какими свойствами обладают слабоформализованные задачи, составляющие область применения искусственного интеллекта?</p> <p>А) Цели этих задач не могут быть представлены в виде математических функций;</p> <p>Б). Алгоритмы достижения этих целей не могут быть описаны строго математически;</p> <p>В) Для решения задач требуются эвристики – утверждения, основанные на опыте, интуиции;</p> <p>Г) Данные и знания, используемые для решения этих задач, могут являться не полными, неоднозначными, ошибочными, противоречивыми, разнородными, динамичными; информация часто может быть представлена в виде изображений, рисунков, знаков, букв, слов, звуков и т.д</p>	<p>ОПК-2.Д.2</p> <p>ОПК-2.Д.4</p>
13	<p>Какие из метрик совместно дают достаточно полную характеристику классификатора при несбалансированной выборке?</p>	<p>ПК-3.Д.1</p> <p>ПК-3.Д.4</p>

	<p>3 Оценка и улучшение качества модели</p> <p>4 Оценка качества модели на тестовой выборке</p>	<p>обучение на обучающей выборке</p> <p>В) Используется для улучшения точности и обобщающей способности модели. Включает в себя анализ ошибок, настройку параметров модели, а также использование регуляризации и ансамблей моделей</p> <p>Г) Этап, на котором проверяется способность модели к обобщению на новые данные, не использовавшиеся в процессе обучения</p> <p>Д) Этап, на котором осуществляется разметка данных</p>	
18	<p>Установите соответствие. Методы оценки качества моделей машинного обучения и их описание</p> <p>1 Кросс-валидация</p> <p>2 Матрица ошибок</p> <p>3 MSE / RMSE</p>	<p>А) Разбиение данных на n частей и последовательное использование каждой части в качестве теста, а оставшихся в качестве обучения</p> <p>Б) Таблица, используемая для оценки точности предсказаний бинарной классификации при несбалансированных классах</p> <p>В) Среднеквадратичная ошибка / корень из среднеквадратичной ошибки для оценки точности регрессионных моделей</p> <p>Г) График, отображающий степень корректности классификации в зависимости от выбранного порогового значения</p>	<p>ОПК-2.Д.2</p> <p>ОПК-2.Д.4</p>
19	<p>Установите соответствие. Типы ансамблей моделей машинного обучения и описание</p> <p>1 Баггинг (Bootstrap Aggregating)</p> <p>2 Бустинг (Boosting)</p> <p>3 Стекинг (Stacking)</p>	<p>А) Метод построения нескольких моделей на основе случайных подвыборок обучающей выборки, агрегирование результатов для получения окончательного прогноза</p> <p>Б) Способ построения последовательности слабых моделей, каждая следующая обучается на данных, на которых предыдущая допускала ошибки</p> <p>В) Метод, в котором несколько моделей обучаются независимо, а затем их результаты используются в качестве признаков в общей модели</p> <p>Г) Метод, при котором несколько моделей голосуют за окончательный прогноз, взвешенное или равномерное количество голосов разных моделей может быть использовано</p>	<p>ОПК-2.Д.2</p> <p>ОПК-2.Д.4</p>
20	<p>Установите соответствие. Примеры алгоритмов машинного</p>		<p>ПК-3.Д.1</p>

	<p>В) Прогнозирование (использование модели);</p> <p>Г) Обучение модели (отбор модели и обучение, выбор метрики качества, перекрёстная проверка, гиперпараметрическая оптимизация).</p>	
24	<p>Укажите последовательность этапов взаимодействия агентов со средой в машинном обучении с подкреплением, которые дали название методу обучения SARSA/</p> <p>А) Состояние; Б) Действие. В) Подкрепление;</p>	<p>ПК-3.Д.1 ПК-3.Д.4</p>
<p>5 тип. Задание открытого типа с развернутым ответом</p> <p>Инструкция: Прочитайте текст и запишите развернутый обоснованный ответ или напишите пропущенное слово/словосочетание</p>		
25	<p>_____ — это процесс, при котором компьютеры могут учиться без явного программирования</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
26	<p>_____ – алгоритм машинного обучения, который находит группы похожих объектов в данных</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
27	<p>Логистическая регрессия, метод опорных векторов (SVM), наивный байесовский классификатор и др. — это модели машинного обучения для решения задачи _____</p>	<p>УК-1.3.2 УК-1.У.1 УК-1.В.2</p>
28	<p>_____ – явление, когда модель слишком точно подстраивается под тренировочные данные, что приводит к плохой обобщающей способности.</p>	<p>ОПК-2.Д.2 ОПК-2.Д.4</p>
29	<p>Функция _____ — это функция, используемая в нейронных сетях для определения значения выходного сигнала нейронов</p>	<p>ОПК-2.Д.2 ОПК-2.Д.4</p>
30	<p>_____ называют эффект противоположный переобучению, когда модель слишком проста, чтобы описать структуру данных.</p>	<p>ПК-3.Д.1 ПК-3.Д.4</p>
31	<p>_____ – это граф, узлами которого являются функции, а ребра связывают функции со своими аргументами.</p>	<p>ПК-3.Д.1 ПК-3.Д.4</p>

Примечание: СИСТЕМА ОЦЕНИВАНИЯ ТЕСТОВЫХ ЗАДАНИЙ.

1-й тип. Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора считается верным, если правильно указана цифра и приведены конкретные аргументы, используемые при выборе ответа.

Полное совпадение с верным ответом – 1 балл.

Неверный ответ или его отсутствие – 0 баллов.

2-й тип. Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора считается верным, если правильно указаны цифры и приведены конкретные аргументы, используемые при выборе ответов.

Полное совпадение с верным ответом – 1 балл.

Если допущены ошибки или ответ отсутствует – 0 баллов.

3-й тип. Задание закрытого типа на установление соответствия считается верным, если установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого столбца).

Полное совпадение с верным ответом – 1 балл.

Неверный ответ или его отсутствие – 0 баллов.

4-й тип. Задание закрытого типа на установление последовательности считается верным, если правильно указана вся последовательность цифр.

Полное совпадение с верным ответом – 1 балл.

Если допущены ошибки или ответ отсутствует – 0 баллов.

5-й тип. Задание открытого типа с развернутым ответом считается верным, если ответ совпадает с эталонным по содержанию и полноте.

Правильный ответ за задание оценивается в 3 балла.

Если допущена одна ошибка\неточность\ответ правильный, но не полный – 1 балл.

Если допущено более 1 ошибки\ответ неправильный\ответ отсутствует – 0 баллов.

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:

Лекция состоит из трёх основных частей: вступительной, основной и заключительной.

Вступительная часть определяет название темы, план и цель лекции. Она призвана заинтересовать и настроить аудиторию. В этой части лекции излагается актуальность, основная идея, связь данной лекции с предыдущими занятиями, ее основные вопросы. Введение должно быть кратким и целенаправленным.

В основной части лекции реализуется научное содержание темы, все главные узловые вопросы, проводится вся система доказательств с использованием наиболее целесообразных методических приёмов. Каждый учебный вопрос заканчивается краткими выводами, логически подводящими студентов к следующему вопросу лекции.

Заключительная часть имеет целью обобщать в кратких формулировках основные идеи лекции, логически завершая её как целостное творение.

Отдельные виды лекций могут иметь свои особенности как по содержанию, так и по структуре.

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

Практические занятия проводятся в компьютерном классе с использованием пакетов программ разработки и отладки программного обеспечения дистрибутива Anaconda.

Anaconda – Дистрибутив Python, предназначенный для крупномасштабной обработки данных, прогнозной аналитики и научных вычислений от компании Continuum Analytics. Это бесплатный, включая коммерческое использование, и готовый к использованию в среде предприятия дистрибутив Python, который объединяет все ключевые библиотеки, необходимые для работы в области науки о данных, математики и разработки. Anaconda уже включает NumPy, SciPy, matplotlib, pandas, IPython, Jupyter Notebook и scikit-learn.

Практические занятия предполагают работу по индивидуальному заданию, связанному с разработкой и отладкой робототехнических систем обработки информации.

По результатам каждого практического задания должен быть подготовлен отчёт, содержащий необходимые теоретические сведения, листинги m-программ, построенные в процессе выполнения работы графические зависимости и т.п. Каждый отчёт должен содержать выводы по проделанной работе и список используемых дополнительных источников.

Каждое практическое занятие завершается собеседованием с преподавателем по представленному отчёту с выставлением оценки.

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.

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.

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.

Контроль качества знаний проводится в форме индивидуального собеседования по материалу отдельных разделов дисциплины, а также проверки отчётов о выполнении практических заданий.

Результаты текущего контроля могут учитываться 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 проводится в форме дифференцированного зачёта.

Дифференцированный зачет – это форма оценки знаний, полученных обучающимся при изучении дисциплины, при выполнении курсовых проектов, курсовых работ, научно-исследовательских работ и прохождении практик с аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Система оценок при проведении промежуточной аттестации осуществляется в соответствии с требованиями Положений «О текущем контроле успеваемости и промежуточной аттестации студентов ГУАП, обучающихся по программам высшего образования» и «О модульно-рейтинговой системе оценки качества учебной работы студентов в ГУАП».

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