

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКАЯ ГОСУДАРСТВЕННАЯ АКАДЕМИЯ»

«УТВЕРЖДАЮ»

Проректор по учебной работе _____ 30 Нагорная
«27» 03 2026



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Анализ данных и машинное обучение

Уровень образовательной программы бакалавриат

Направление подготовки 09.03.04 Программная инженерия

Направленность (профиль) Программная инженерия

Форма обучения очная

Срок освоения ОП 4 года

Институт Цифровых технологий

Кафедра разработчик РПД Цифровая инженерия и сетевые технологии

Выпускающая кафедра Цифровая инженерия и сетевые технологии

Начальник учебно-методического управления _____ Семенова Л.У.

Директор института ЦТ _____ Кумратова А.М.

Заведующий выпускающей кафедрой _____ Эльканова Л.М.

г. Черкесск, 2026 г.

СОДЕРЖАНИЕ

- 1. Цели освоения дисциплины**
 - 2. Место дисциплины в структуре образовательной программы**
 - 3. Планируемые результаты обучения по дисциплине**
 - 4. Структура и содержание дисциплины**
 - 4.1. Объем дисциплины и виды учебной работы
 - 4.2. Содержание дисциплины
 - 4.2.1. Разделы (темы) дисциплины, виды учебной деятельности и формы контроля
 - 4.2.2. Лекционный курс
 - 4.2.3. Лабораторный практикум
 - 4.2.4. Практические занятия
 - 4.3. Самостоятельная работа обучающегося
 - 5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**
 - 6. Образовательные технологии**
 - 7. Учебно-методическое и информационное обеспечение дисциплины**
 - 7.1. Перечень основной и дополнительной учебной литературы
 - 7.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»
 - 7.3. Информационные технологии, лицензионное программное обеспечение
 - 8. Материально-техническое обеспечение дисциплины**
 - 8.1. Требования к аудиториям (помещениям, местам) для проведения занятий
 - 8.2. Требования к оборудованию рабочих мест преподавателя и обучающихся
 - 8.3. Требования к специализированному оборудованию
 - 9. Особенности реализации дисциплины для инвалидов и лиц с ограниченными возможностями здоровья**
- Приложение 1. Фонд оценочных средств**

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Анализ данных и машинное обучение» являются: формирование у обучающихся понимания современных методов и инструментальных средств применения машинного обучения для анализа данных.

При этом задачами дисциплины являются:

- знать типы задач, модели и алгоритмы машинного обучения, особенности нейронных сетей;
- выбирать современные инструментальные средства, необходимые библиотеки и пакеты для анализа данных;
- применять методы и алгоритмы, используемые в сфере машинного обучения для разработки приложений на языке Python.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

2.1. Дисциплина «Анализ данных и машинное обучение» относится к части, формируемой участниками образовательных отношений Блока 1. Дисциплины (модули), имеет тесную связь с другими дисциплинами.

2.2. В таблице приведены предшествующие и последующие дисциплины, направленные на формирование компетенций дисциплины в соответствии с матрицей компетенций ОП.

Предшествующие дисциплины, направленные на формирование компетенций

№ п/п	Предшествующие дисциплины	Последующие дисциплины
1	Алгоритмы и структуры данных Математическое моделирование Математические основы искусственного интеллекта Логическое и функциональное программирование	Системы искусственного интеллекта Интеллектуальные системы анализа Информационная безопасность и защита информации

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Планируемые результаты освоения образовательной программы (ОП) – компетенции обучающихся определяются требованиями стандарта по направлению подготовки 09.03.04 Программная инженерия и формируются в соответствии с матрицей компетенций ОП

№ п/п	Номер/ Индекс компетенции	Наименование Компетенции (или ее части)	В результате изучения дисциплины обучающиеся должны:
1	2	3	4
1.	ПК-1	Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности	ПК-1.1. Выбирает современные инструментальные средства для проектирования и конструирования программного обеспечения ПК-1.3. Осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий ПК-1.4. Применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах
	ПК-2	Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности	ПК-2.1. Разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения. ПК-2.4. Оценивает временную и емкостную сложность программного обеспечения ПК2.6. Работает с современными системами программирования и проектирования

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Вид учебной работы	Всего часов	Семестр	Семестр	
		№7	№8	
		Часов	Часов	
1	2	3		
Аудиторная контактная работа(всего)	102	42	60	
В том числе:				
Лекции(Л)	34	14	20	
Практические занятия(ПЗ), Семинары(С)	-	-	-	
Лабораторные работы(ЛР)	68	28	40	
Контактная внеаудиторная работа, в том числе:	3,7	1,7	2	
Групповые и индивидуальные консультации	3,7	1,7	2	
Самостоятельная работа обучающегося(СРО)(всего)	128	64	64	
Работа с книжными источниками	22	14	14	
Подготовка к лабораторным занятиям	23	12	16	
Выполнение контрольных работ	26	24	12	
Работа с электронными источниками	26	6	6	
Подготовка к тестированию	22	8	16	
Промежуточная аттестация	Зачет (З)	3	3	-
	Прием зач., час.	0,3	0,3	
	экзамен (Э) в том числе:	Э(18)	-	Э(18)
	Прием экз., час.	0,5	-	0,5
	Консультация, час.	2	-	2
	СРО, час.	15,5	-	15,5
ИТОГО:	Часов	252	108	144
Общая трудоемкость	зачетных единиц	7	3	4

СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.2.1. Разделы(темы) дисциплины, виды учебной деятельности и формы контроля

№ п/п	Наименование раздела дисциплины	Виды учебной деятельности, включая самостоятельную работу обучающихся (в часах)					Формы текущей и промежуточной аттестации)
		Л	ЛР	ПЗ	СРО	Всего	
1	2	3	4	5	6	7	8
Семестр 7							
1.	Раздел 1. Введение в анализ данных и машинное обучение	2	4	-	20	26	лабораторная работа, контрольная работа
2.	Раздел 2. Основы Python для анализа данных	4	6	-	22	32	лабораторная работа, контрольная работа
3	Раздел 3. Задачи машинного обучения	8	18	-	22	48	Текущий тестовый контроль, лабораторная работа, контрольная работа
4.	Контактная внеаудиторная работа					1,7	Индивидуальные и групповые консультации
5.	Промежуточная аттестация					0,3	зачет
Итого часов в 7 семестре:		14	28	-	64	108	

№ п/п	Наименование раздела дисциплины	Виды учебной деятельности, включая самостоятельную работу обучающихся (в часах)					Формы текущей и промежуточной аттестации)
		Л	ЛР	ПЗ	СРО	Всего	
1	2	3	4	5	6	7	8
Семестр 8							
1.	Раздел 1. Задачи машинного обучения.	10	18	-	16	44	лабораторная работа, контрольная работа
2.	Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	6	10	-	16	32	лабораторная работа, контрольная работа
3	Раздел 3. Введение в нейронные сети	10	12		16	38	Текущий тестовый контроль, лабораторная работа, контрольная работа
4.	Контактная внеаудиторная работа					2	Индивидуальные и групповые консультации
5.	Промежуточная аттестация					18	экзамен
Итого часов в 8 семестре:		20	40	-	64	144	

4.2.2. Лекционный курс

№ п/п	Наименование Раздела дисциплины	Наименование темы лекции	Содержание лекции	Всего часов
Семестр 7				
1	2	3	4	5
1.	Раздел 1. Введение в анализ данных машинное обучение	Тема 1.1 Введение в анализ данных и машинное обучение	Большие данные. Характеристики больших данных. Инструменты работы с большими данными. Машинное обучение как область ИИ. Процесс построения модели ML.	2

2.	Раздел 2. Основы Python для анализа данных	Тема 2.1 Основы Python для анализа данных.	Python– интерпретируемый язык программирования. Синтаксис.Интерактивный режим,интерпретатор. Создание скриптов. Типы данных: целые числа, числа с плавающей точкой,строки. Переменные Python. Конструкции if— заголовок, в котором проверяется условие выполнения строк кода после двоеточия. Сложная форма ветвления if— else. Инструкция if-elif-else.	2
		Тема 2.2 Списки, кортежи, словари Цикл for в языке программирования Python. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.	Списки, кортежи, словари Цикл for-циклобхода, заданного множества элементов и выполнения в своем теле различных операций над ними. Инструкция def - команда языка программирования Python, позволяющая создавать функцию.	2
3	Раздел 3. Задачи машинного обучения	Тема 3.1 Задачи машинного обучения с учителем	Задачи машинного обучения с учителем, задача классификации, задача регрессии, метрики, алгоритмы, методы	4
		Тема 3.2 Задачи машинного обучения без учителя и предварительная обработка данных	Задачи машинного обучения без учителя, кластеризация, меры расстояния, алгоритмы кластеризации, предварительная обработка данных	2

		Тема 3.3 Задачи машинного обучения. Понижение размерности	Задача понижения размерности Алгоритмы отбора признаков Метод фильтрации признаков Оберточные методы отбора признаков Встроенные методы отбора Алгоритмы выделения новых признаков Задача определения тональности	2
Итого часов в 7 семестре:				14

№ п/п	Наименование раздела дисциплины	Наименование темы лекции	Содержание лекции	Всего часов
Семестр 8				
1	2	3	4	5
1.	Раздел 1. Задачи машинного обучения.	Тема 1.1 Задачи машинного обучения.	Понятие «ассоциативные правила» Меры для определения ассоциаций Формальное определение (ассоциативное правило) Алгоритмы ассоциативных правил Рекомендательные системы	2
		Тема 1.2 Обучение с подкреплением	Обучение с подкреплением Стратегия действий агента при Q-обучении Подход на основе Policy Gradient. Подход на основе Q-learning. Марковский процесс принятия решений	2
		Тема 1.3 Алгоритмы машинного обучения	Решающие деревья Метод опорных векторов Наивный байесовский алгоритм Кросс-валидация	4
		Тема 1.4 Ансамблирование	Ансамбли моделей Бэггинг Бустинг Стекинг	2

2.	Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	Тема 2.1 Введение Pandas. Библиотека Pandas и библиотеки визуализации	Области применения: Data Science, аналитика, машинное обучение, бизнес-отчётность. Преимущества перед Excel/SQL Основные структуры данных . Series. DataFrame. Загрузка и сохранение данных. Форматы данных: CSV, Excel, JSON, SQL. Чтение файлов. Визуализация данных (20 мин) Интеграция с Matplotlib: Библиотека Seaborn:	2
		Тема 2.2 Введение в scikit-Learn	Scikit-Learn для обучения с учителем, для обучения без учителя Работа с моделями Встроенные датасеты Разбиения выборок: cross_validation Качество: metrics	2
		Тема 2.3 Введение Numpy	Основные понятия: массив ndarray Создание массивов Индексация и срезы Векторизованные операции Изменение формы массивов Стекинг и разделение Агрегация и статистика Работа с типами данных и памятью	2
4	Раздел 3. Введение в нейронные сети	Тема 3.1 Введение в нейронные сети	Искусственный нейрон. Архитектура полносвязной нейронной сети. Функции активации. Обзор фреймворков (Keras/TensorFlow). Сверточные и рекуррентные нейронные сети. Обучение нейронной сети (Forward/Backward propagation).	4
Итого часов в 8 семестре:				20
Всего часов:				20

4.2.3. Лабораторный практикум

№ п/п	Наименование раздела дисциплины	Наименование лабораторного занятия	Содержание лабораторного занятия	Всего часов
Семестр 7				
1	2	3	4	5
1.	Раздел 1. Введение в анализ данных и машинное обучение	Тема 1.1 Введение в анализ данных и машинное обучение	Большие данные. Характеристики больших данных. Инструменты работы с большими данными. Машинное обучение как область ИИ. Процесс построения модели ML	4
2.	Раздел 2. Основы Python для анализа данных	Тема 2.1 Основы Python для анализа данных.	Python–интерпретируемый язык программирования. Синтаксис. Интерактивный режим, интерпретатор. Создание скриптов. Типы данных: целые числа, числа с плавающей точкой, строки. Переменные Python. Конструкции if–заголовок, в котором проверяется условие выполнения строк кода после двоеточия. Сложная форма ветвления if– else. Инструкция if-elif-else.	2
		Тема 2.2 Списки, кортежи, словари Цикл for в языке программирования Python. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.	Списки, кортежи, словари Цикл for-цикл обхода, заданного множества элементов и выполнения в своем теле различных операций над ними. Инструкция def - команда языка программирования Python, позволяющая создавать функцию.	4

3	Раздел 3. Задачи машинного обучения	Тема 3.1 Задачи машинного обучения с учителем	Задачи машинного обучения с учителем, задача классификации, задача регрессии, метрики, алгоритмы, методы	6
		Тема 3.2 Задачи машинного обучения без учителя и предварительная обработка данных	Задачи машинного обучения без учителя, кластеризация, меры расстояния, алгоритмы кластеризации, предварительная обработка данных	6
		Тема 3.3 Задачи машинного обучения. Понижение размерности	Задачи машинного обучения. Понижение размерности, отбор и выделение признаков. Методы отбора признаком, алгоритмы выделения признаков	6
Итого часов в 7 семестре:				14

№п /п	Наименование раздела дисциплины	Наименование лабораторного занятия	Содержание Лабораторного занятия	Всего часов
Семестр 8				
1	2	3	4	5
1.	Раздел 1. Задачи машинного обучения.	Тема 1.1 Ассоциативные правила.	Понятие «ассоциативные правила» Меры для определения ассоциаций Формальное определение (ассоциативное правило) Алгоритмы ассоциативных правил Рекомендательные системы	4
		Тема 1.2 Обучение с подкреплением	Обучение с подкреплением Стратегия действий агента при Q-обучении Подход на основе Policy Gradient. Подход на основе Q-learning. Марковский процесс принятия решений	4
		Тема 1.3 Алгоритмы машинного обучения	Решающие деревья Метод опорных векторов Наивный байесовский алгоритм Кросс-валидация	4

		Тема 1.4 Ансамблирование	Ансамбли моделей Бэггинг Бустинг Стекинг	2
2.	Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	Тема 2.1 Введение Pandas. Библиотека Pandas и библиотеки визуализации	Работа в Pandas Загрузить CSV-файл с данными о пассажирах Титаника. Вывести общую информацию о наборе данных. Отфильтровать пассажиров старше 40 лет. Построить гистограмму распределения возрастов. Сохранить очищенные данные в новый файл.	4
		Тема 2.2 Введение в scikit-Learn	Scikit-Learn для обучения с учителем, для обучения без учителя Работа с моделями Встроенные датасеты Разбиения выборок: cross_validation Качество: metrics	2
		Тема 2.3 Введение NumPy	Создать массив 5×5 со случайными числами. Найти сумму каждой строки и столбца. Заменить все значения > 0.5 на 1, остальные — на 0. Трансформировать массив в 1D и отсортировать. Вычислить среднее и стандартное отклонение полученного вектора.	4
3	Раздел 3. Введение в нейронные сети	Тема 3.1 Введение в нейронные сети	Искусственный нейрон. Архитектура полносвязной нейронной сети. Функции активации. Обзор фреймворков (Keras/TensorFlow). Сверточные и рекуррентные нейронные сети. Обучение нейронной сети (Forward/Backward propagation).	12

			Изучение простейших моделей нейронной сети на примере библиотеки Sklearn. Изучение алгоритма обратного распространения ошибки. Изучение общих рекомендаций по выполнению экспериментов Изучение общих рекомендаций по выполнению экспериментов	
Итого часов в 8 семестре:				40
Всего часов:				40

4.2.4. Практические занятия не предполагаются

4.3. САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ

№ п/п	Наименование раздела (темы) дисциплины	№ п/п	Виды СРО	Всего часов
1	2	3	4	5
Семестр 7				
1.	Раздел 1. Введение в анализ данных и машинное обучение	1.1.	Работа с книжными источниками	8
		1.2.	Выполнение контрольных работ	8
		1.3.	Подготовка к лабораторным занятиям	4
2.	Раздел 2. Основы Python для анализа данных	2.1.	Работа с электронными источниками	6
		2.2.	Выполнение контрольных работ	8
		2.3.	Подготовка к тестированию	8
3.	Раздел 3. Задачи машинного обучения	3.1.	Работа с книжными источниками	6
		3.2.	Выполнение контрольных работ	8
		3.3.	Подготовка к лабораторным занятиям	8
Итого часов в 7 семестре:				64

№ п/п	Наименование раздела (темы) дисциплины	№ п/п	Виды СРО	Всего часов
1	2	3	4	5
Семестр 8				
1.	Раздел 1. Задачи машинного обучения	1.1.	Работа с книжными источниками	6
		1.2.	Выполнение контрольных работ	6
		1.3.	Подготовка к лабораторным занятиям	8
2.	Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	2.1.	Работа с электронными источниками	6
		2.2.	Подготовка к тестированию	8

		2.3	Выполнение контрольных работ	6
3.	Раздел 4. Введение в нейронные сети	3.1.	Работа с книжными источниками	8
		3.2.	Подготовка к тестированию	8
		3.3.	Подготовка к лабораторным занятиям	8
Итого часов в 8 семестре:				64
Всего часов:				128

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

5.1. Методические указания для подготовки обучающихся к лекционным занятиям

Лекция является основной формой обучения в высшем учебном заведении. Записи лекций в конспектах должны быть избирательными, полностью следует записывать только определения. В конспекте рекомендуется применять сокращение слов, что ускоряет запись. Вопросы, возникающие в ходе лекции, рекомендуется записывать на полях и после окончания лекции обратиться за разъяснением к преподавателю.

Работа над конспектом лекции осуществляется по этапам:

- повторить изученный материал по конспекту;
- непонятные положения отметить на полях и уточнить;
- неоконченные фразы, пропущенные слова и другие недочеты в записях устранить, пользуясь материалами из учебника и других источников;
- завершить техническое оформление конспекта (подчеркивания, выделение главного, выделение разделов, подразделов и т.п.).

Самостоятельную работу следует начинать с доработки конспекта, желательно в тот же день, пока время не стерло содержание лекции из памяти. Работа над конспектом не должна заканчиваться с прослушивания лекции. После лекции, в процессе самостоятельной работы, перед тем, как открыть тетрадь с конспектом, полезно мысленно восстановить в памяти содержание лекции, вспомнив ее структуру, основные положения и выводы.

С целью доработки необходимо прочитать записи, восстановить текст в памяти, а также исправить опiski, расшифровать не принятые ранее сокращения, заполнить пропущенные места, понять текст, вникнуть в его смысл. Далее прочитать материал по рекомендуемой литературе, разрешая в ходе чтения, возникшие ранее затруднения, вопросы, а также дополнения и исправляя свои записи. Записи должны быть наглядными, для чего следует применять различные способы выделений. В ходе доработки конспекта углубляются, расширяются и закрепляются знания, а также дополняется, исправляется и совершенствуется конспект. Еще лучше, если вы переработаете конспект, дадите его в новой систематизации записей. Это, несомненно, займет некоторое время, но материал вами будет хорошо проработан, а конспективная запись его приведена в удобный для

запоминания вид. Введение заголовков, скобок, обобщающих знаков может значительно повысить качество записи. Этому может служить также подчеркивание отдельных мест конспекта красным карандашом, приведение на полях или на обратной стороне листа краткой схемы конспекта и др.

Подготовленный конспект и рекомендуемая литература используется при подготовке к практическому занятию. Подготовка сводится к внимательному прочтению учебного материала, к выводу с карандашом в руках всех утверждений и формул, к решению примеров, задач, к ответам на вопросы, предложенные в конце лекции преподавателем или помещенные в рекомендуемой литературе. Примеры, задачи, вопросы по теме являются средством самоконтроля.

Непрерывным условием глубокого усвоения учебного материала является знание основ, на которых строится изложение материала. Обычно преподаватель напоминает, какой ранее изученный материал и в какой степени требуется подготовить к очередному занятию. Эта рекомендация, как и требование систематической и серьезной работы над всем лекционным курсом, подлежит безусловному выполнению. Потери логической связи как внутри темы, так и между ними приводит к негативным последствиям: материал учебной дисциплины перестает основательно восприниматься, а творческий труд подменяется утомленным переписыванием. Обращение к ранее изученному материалу не только помогает восстановить в памяти известные положения, выводы, но и приводит разрозненные знания в систему, углубляет и расширяет их. Каждый возврат к старому материалу позволяет найти в нем что-то новое, переосмыслить его с иных позиций, определить для него наиболее подходящее место в уже имеющейся системе знаний. Неоднократное обращение к пройденному материалу является наиболее рациональной формой приобретения и закрепления знаний. Очень полезным, но, к сожалению, еще мало используемым в практике самостоятельной работы, является предварительное ознакомление с учебным материалом. Даже краткое, беглое знакомство с материалом очередной лекции дает многое. Обучающиеся получают общее представление о ее содержании и структуре, о главных и второстепенных вопросах, о терминах и определениях. Все это облегчает работу на лекции и делает ее целеустремленной.

5.2. Методические указания для подготовки обучающихся к лабораторным занятиям

Ведущей дидактической целью лабораторных занятий является систематизация и обобщение знаний по изучаемой теме, приобретение практических навыков по тому или другому разделу курса, закрепление полученных теоретических знаний. Лабораторные работы сопровождают и поддерживают лекционный курс. Подготовка к лабораторным занятиям и практикумам носит различный характер, как по содержанию, так и по сложности исполнения.

Многие лабораторные занятия требуют большой исследовательской работы, изучения дополнительной научной литературы. Прежде чем приступить к выполнению такой работы, обучающемуся необходимо ознакомиться обстоятельно с содержанием задания, уяснить его, оценить с точки зрения восприятия и запоминания все составляющие его компоненты. Это очень важно, так как при проработке соответствующего материала по конспекту лекции или по рекомендованной литературе могут встретиться определения, факты, пояснения, которые не относятся непосредственно к заданию. Обучающийся должен хорошо знать и понимать содержание задания, чтобы быстро оценить и отобрать нужное из читаемого. Далее, в соответствии со списком рекомендованной литературы, необходимо отыскать материал к данному заданию по всем пособиям.

Весь подобранный материал нужно хотя бы один раз прочитать или внимательно просмотреть полностью. По ходу чтения помечаются те места, в которых содержится ответ на вопрос, сформулированный в задании. Читая литературу по теме, обучающийся должен мысленно спрашивать себя, на какой вопрос задания отвечает тот или иной абзац прорабатываемого пособия. После того, как материал для ответов подобран, желательно хотя бы мысленно, а лучше всего устно или же письменно, ответить на все вопросы. В случае если обнаружится пробел в знаниях, необходимо вновь обратиться к литературным источникам и проработать соответствующий раздел. Только после того, как преподаватель убедится, что обучающийся хорошо знает необходимый теоретический материал, что его ответы достаточно аргументированы и доказательны, можно считать обучающегося подготовленным к выполнению лабораторных работ.

5.3. Методические указания для подготовки обучающихся к практическим занятиям

Учебным планом не предусмотрено

5.4. Методические указания для подготовки обучающихся к контрольным работам

В процессе подготовки и проведения занятий обучающиеся закрепляют полученные ранее теоретические знания, приобретают навыки их практического применения, опыт рациональной организации учебной работы, готовятся к сдаче экзамена.

Поскольку активность на практических занятиях является предметом внутрисеместрового контроля его продвижения в освоении курса, подготовка к таким занятиям требует ответственного отношения.

При подготовке к занятию в первую очередь должны использовать материал лекций и соответствующих литературных источников. Самоконтроль качества подготовки к каждому занятию осуществляют, проверяя свои знания и отвечая на вопросы для самопроверки по соответствующей теме.

Входной контроль осуществляется преподавателем в виде проверки и актуализации знаний обучающихся по соответствующей теме.

Выходной контроль осуществляется преподавателем проверкой качества и полноты выполнения задания.

Подготовку к индивидуальным заданиям каждый обучающийся должен начать с ознакомления с планом задания, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала, а затем изучение обязательной и дополнительной литературы, рекомендованной к данной теме.

Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности обучающегося свободно ответить на теоретические вопросы, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий.

Предлагается следующая опорная схема подготовки к индивидуальным заданиям.

1. Ознакомление с темой занятия. Выделение главного (основной темы) и второстепенного (подразделы, частные вопросы темы).

2. Освоение теоретического материала по теме с опорой на лекционный материал, учебник и другие учебные ресурсы. Самопроверка: постановка вопросов, затрагивающих основные термины, определения и положения по теме, и ответы на них.

3. Выполнение практического задания. Обнаружение основных трудностей, их решение с помощью дополнительных интеллектуальных усилий и/или подключения дополнительных источников информации.

4. Решение типовых заданий лабораторной работы.

Обучающийся при подготовке к индивидуальному заданию может консультироваться с преподавателем и получать от него наводящие разъяснения, задания для самостоятельной работы.

Дидактические цели занятия: углубление, систематизация и закрепление знаний, превращение их в убеждения; проверка знаний; привитие умений и навыков самостоятельной работы с книгой; развитие культуры речи, формирование умения аргументировано отстаивать свою

точку зрения, отвечать на вопросы слушателей; умение слушать других, задавать вопросы. Задачи: стимулировать регулярное изучение программного материала, первоисточников; закреплять знания, полученные на уроке и во время самостоятельной работы; обогащать знаниями благодаря выступлениям товарищей и учителя на занятии, корректировать ранее полученные знания.

Функции практического занятия:

- учебная (углубление, конкретизация, систематизацию знаний, усвоенных во время занятий и в процессе самостоятельной подготовки к семинару);
- развивающая (развитие логического мышления учащихся обучающихся, приобретение ими умений работать с различными литературными источниками, формирование умений и навыков анализа фактов, явлений, проблем и т.д.);
- воспитательная (воспитание ответственности, работоспособности, воспитание культуры общения и мышления, привитие интереса к изучению предмета, формирование потребности рационализации и учебно-познавательной деятельности и организации досуга)
- диагностическая -коррекционную и контролирующую (контроль за качеством усвоения обучающимися учебного материала, выявление пробелов в его усвоении и их преодоления)
- организация самостоятельной работы обучающихся содержит объяснение содержания задачи, методики его выполнения, краткую аннотацию рекомендованных источников информации, предложения по выполнению индивидуальных заданий.

5.5. Методические указания по самостоятельной работе обучающихся

Работа с литературными источниками и интернет ресурсами

В процессе подготовки к лабораторным занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы.

Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме.

Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Тестирование

Тестирование представляет собой средство контроля усвоения учебного материала темы или раздела дисциплины. При самостоятельной подготовке к тестированию обучающемуся необходимо:

а) проработать информационный материал по дисциплине, проконсультироваться с преподавателем по вопросу выбора учебной литературы;

б) выяснить все условия тестирования заранее, узнать, сколько тестов будет предложено, сколько времени отводится на тестирование, какова система оценки результатов и т.д.

в) приступая к работе с тестами, внимательно и до конца прочитать вопрос и предлагаемые варианты ответов, выбрать правильные (их может быть несколько).

г) в процессе решения желательно применять несколько подходов в решении задания, это позволяет максимально гибко оперировать методами решения, находя каждый раз оптимальный вариант.

д) если встретился чрезвычайно трудный вопрос, не тратить много времени на него, перейти к другим тестам, вернуться к трудному вопросу в конце.

е) обязательно оставить время для проверки ответов, чтобы избежать механических ошибок.

– аргументировать.

Работа с книжными и электронными источниками

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

Методические рекомендации для подготовки к тестовому контролю

Тесты - это задания, предусматривающие конкретный, краткий, четкий ответ на имеющиеся эталоны ответов. Готовясь к тестированию, необходимо проработать информационный материал по дисциплине.

Обучающемуся необходимо проконсультироваться с преподавателем по вопросу выбора учебной литературы; четко выясните все условия тестирования заранее.

Приступая к работе с тестами, внимательно и до конца необходимо прочитать вопрос и предлагаемые варианты ответов. Выбираем правильные ответы, или дополняем. В процессе решения желательно применять несколько подходов в решении задания. Это позволяет максимально гибко оперировать методами решения, находя каждый раз оптимальный вариант.

Промежуточная аттестация

По итогам 7 семестра проводится зачет, 8 семестра проводится экзамен. При подготовке к сдаче зачета, экзамена рекомендуется пользоваться материалами лабораторных занятий и материалами, изученными в ходе текущей самостоятельной работы.

Зачет, экзамен проводится в устной форме, включает подготовку и ответы обучающегося на теоретические вопросы. По итогам экзамена выставляется оценка.

По итогам обучения проводится экзамен, к которому допускаются студенты, имеющие положительные результаты по защите лабораторных работ.

6. Образовательные технологии

№ п/п	Виды учебной работы	Образовательные технологии	Всего часов
			ОФО
1	2	3	4
Семестр 7			
1	Лекция. Введение в анализ данных и машинное обучение	Лекция –презентация	2
2	Лекция. Основы Python для анализа данных.	Лекция –презентация	2
3	Лекция Задачи машинного обучения с учителем	Диалоговые технологии	2
4	Лекция Задачи машинного обучения без учителя и предварительная обработка данных	Диалоговые технологии	4
5	Лабораторная работа Задачи машинного обучения без учителя и предварительная обработка данных	Технология развития критического мышления	2
Итого часов в 7 семестре:			10

№ п/п	Виды учебной работы	Образовательные технологии	Всего часов
			ОФО
1	2	3	4
Семестр 8			
1	Лекция Задачи машинного обучения	Лекция –презентация	4

2	Лекция Введение Pandas. Библиотека Pandas и библиотеки визуализации	Лекция –презентация	2
3	Лекция Введение NumPy.	Диалоговые технологии	2
5	Лабораторная работа Работа с простейшими моделями нейронных сетей	Технология развития критического мышления	4
Итого часов в 8 семестре:			10

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

7.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. — Ульяновск : Ульяновский государственный технический университет, 2017. — 291 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/106120.html>.
2. Воронова, Л. И. MachineLearning: регрессионные методы интеллектуального анализа данных : учебное пособие / Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2018. — 82 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/81325.html>
3. Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Замятин, А. В. Введение в интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Томск : Издательский Дом Томского государственного университета, 2016. — 119 с. — ISBN 978-5-94621-531-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/109021.html>.
4. Жуковский, О. И. Информационные технологии и анализ данных : учебное пособие / О. И. Жуковский. — Томск : Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. — 130 с. — ISBN 978-5-4332-0158-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/72106.html>.
5. Томск : Издательский Дом Томского государственного университета, 2020. — 194 с. — ISBN 978-5-94621-898-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/116889.html>.
6. Целых, А. Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» / А. Н. Целых, А. А. Целых, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2021. — 130 с. — ISBN 978-5-9275-3783-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/117165.html>.
7. Целых А.Н. Извлечение знаний методами машинного обучения [Электронный ресурс]: учебное пособие по курсам «Модели и методы инженерии знаний», «Методы машинного обучения»/ Целых А.Н., Котов Э.М.— Электрон. текстовые данные.— Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2022.— 105 с.— Режим доступа: <https://ipr-smart.ru/131448> .— IPR SMART, по паролю
8. Целых, А. Н. Принятие решений на основе методов машинного обучения : учебное пособие по курсам «Модели и методы инженерии знаний», «Методы анализа больших данных» / А. Н. Целых, Н. В. Драгныш, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2022. — 113 с. — ISBN 978-5-9275-4246-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/131458.html>.

Дополнительная литература

1. Запечников, С. В. Основы интеллектуального анализа данных и машинного обучения: конспект лекций : учебное пособие / С. В. Запечников. — Москва : Национальный исследовательский ядерный университет «МИФИ», 2022. — 136 с. — ISBN 978-5-7262-2856-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/132687.html>.
2. Обработка данных магниторазведки: обратные задачи и машинное обучение : учебное пособие / Д. С. Киселев, А. М. Грифф, М. Г. Персова, Ю. Г. Соловейчик. —

- Новосибирск : Новосибирский государственный технический университет, 2021. — 76 с. — ISBN 978-5-7782-4556-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/126576.html>.
3. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/89482.html>.
4. Сараев, П. В. Методы машинного обучения : методические указания и задания к лабораторным работам по курсу / П. В. Сараев. — Липецк : Липецкий государственный технический университет, ЭБС АСВ, 2017. — 48 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/83183.html>.

Методическая литература

нет

7.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

<https://intuit.ru/> Сайт национального открытого университета ИНТУИТ

Google Colab (**Google Colaboratory**) — бесплатная облачная платформа от Google, которая позволяет писать и выполнять код на языке Python прямо в браузере.

7.3. Информационные технологии, лицензионное программное обеспечение.

Лицензионное программное обеспечение	Реквизиты лицензий/ договоров
Антивирус Dr.Web Desktop Security Suite	Лицензионный договор № 621 Срок действия: с 25.09.2025 до 24.09.2026
Цифровой образовательный ресурс IPR SMART	Лицензионный договор № 12873/25П от 02.07.2025 г. Срок действия: с 01.07.2025 г. до 30.06.2026 г.
ЛИРА	Сублицензионный договор № 2066/А от 21.01.2014 г.
Бесплатное ПО	
LibreOffice, OpenOffice, МойОфис, Sumatra PDF, 7-Zip, Adobe Acrobat Reader, Python, Google Colab	

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1. Требования к аудиториям (помещениям, местам) для проведения занятий

1. Учебная аудитория для проведения занятий лекционного типа

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 10 шт., парты - 7 шт., стулья - 24 шт., доска меловая - 1 шт.

Лабораторное оборудование, технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Персональный компьютер – 10 шт.

Экран настенный рулонный – 1 шт.

2. Лаборатория компьютерной графики

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 10 шт., парты - 7 шт., стулья - 24 шт., доска меловая - 1 шт.

Лабораторное оборудование, технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Персональный компьютер – 10 шт.

Экран настенный рулонный – 1 шт.

3. Учебная аудитория для проведения занятий семинарского типа, курсового проектирования (выполнение курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации

Специализированная мебель:

Доска меловая - 1 шт., парты - 35 шт., стулья - 66 шт., кафедра настольная - 1 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Настенный экран - 1 шт.

Проектор - 1 шт.

Ноутбук – 1 шт.

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 13 шт., стул мягкий - 1 шт., парты - 7 шт., стулья - 21 шт., доска меловая - 1 шт.

Технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Сист. блок iRu Ergo-Corp 1294 E8400(3000)1024 80 DVD-RW FDD K – 1 шт.;

Сист. Блок Intel P Dual E5700/HDD 500Gb/DDR3 2Gb – 9 шт.;

Монитор LCD e`machines LCD 18.5' – 8 шт.;

Монитор LOC 18.5' – 2 шт.;

Экран настенный рулонный 200x200 – 1 шт.

4. Помещение для самостоятельной работы. Библиотечно-издательский центр.

Отдел обслуживания печатными изданиями

Специализированная мебель:

Рабочие столы на 1 место – 21 шт.

Стулья – 55 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Экран настенный – 1 шт.

Проектор – 1 шт.

Ноутбук – 1 шт.

Информационно-библиографический отдел.

Специализированная мебель:

Рабочие столы на 1 место - 6 шт.

Стулья - 6 шт.

Компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду ФГБОУ ВО «СевКавГА»:

Персональный компьютер – 1 шт.

Сканер – 1 шт.

МФУ – 1 шт.

Отдел обслуживания электронными изданиями

Специализированная мебель:

Рабочие столы на 1 место – 24 шт.

Стулья – 24 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Интерактивная система - 1 шт.

Монитор – 21 шт.

Сетевой терминал -18 шт.

Персональный компьютер -3 шт.

МФУ – 2 шт.

Принтер –1шт.

8.2. Требования к оборудованию рабочих мест преподавателя и обучающихся

- 1 Рабочее место преподавателя, оснащенное компьютером с доступом в сеть «Интернет».
- 2 Рабочие места обучающихся, оснащенные компьютерами с доступом в сеть «Интернет», предназначенные для работы в электронной образовательной среде.

8.3. Требования к специализированному оборудованию

- нет

9. ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Для обеспечения образования инвалидов и обучающихся с ограниченными возможностями здоровья разрабатывается (в случае необходимости) адаптированная образовательная программа, индивидуальный учебный план с учетом особенностей их психофизического развития и состояния здоровья, в частности применяется индивидуальный подход к освоению дисциплины, индивидуальные задания: рефераты, письменные работы и, наоборот, только устные ответы и диалоги, индивидуальные консультации, использование диктофона и других записывающих средств для воспроизведения лекционного и семинарского материала.

В целях обеспечения обучающихся инвалидов и лиц с ограниченными возможностями здоровья комплектуется фонд основной учебной литературой, адаптированной к ограничению электронных образовательных ресурсов, доступ к которым организован в БИЦ Академии. В библиотеке проводятся индивидуальные консультации для данной категории пользователей, оказывается помощь в регистрации и использовании сетевых и локальных электронных образовательных ресурсов, предоставляются места в читальном зале

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ПО ДИСЦИПЛИНЕ «Анализ данных и машинное
обучение»

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

«Анализ данных и машинное обучение»

Компетенции, формируемые в процессе изучения дисциплины

Индекс	Формулировка компетенции
ПК-1	Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности
ПК-2	Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности

2. Этапы формирования компетенции в процессе освоения дисциплины

Основными этапами формирования указанных компетенций при изучении обучающимися дисциплины являются последовательное изучение содержательно связанных между собой разделов (тем) учебных занятий. Изучение каждого раздела (темы) предполагает овладение обучающимися необходимыми компетенциями. Результат аттестации обучающихся на различных этапах формирования компетенций показывает уровень освоения компетенций студентами.

Этапность формирования компетенций прямо связана с местом дисциплины в образовательной программе.

Разделы (темы) дисциплины	Формируемые компетенции (коды)	
	ПК-1	ПК-2
7 семестр	2	3
Раздел 1. Введение в анализ данных и машинное обучение	+	+
Раздел 2. Основы Python для анализа данных	+	+
Раздел 3. Задачи машинного обучения	+	+
Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	+	+
8 семестр		
Раздел 1. Задачи машинного обучения.	+	+
Раздел 2. Основные библиотеки и инструменты Python для анализа больших данных	+	+

Раздел 3. Введение в нейронные сети		
-------------------------------------	--	--

1. Показатели, критерии и средства оценивания компетенций, формируемых в процессе изучения дисциплины

ПК-1 Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности

Индикаторы достижения компетенции	Критерии оценивания результатов обучения				Средства оценивания результатов обучения	
	неудовлетв	удовлетв	хорошо	отлично	Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6	7
ПК-1.1. Выбирает современные инструментальные средства для проектирования и конструирования программного обеспечения	Не знает основные принципы применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Неполные представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Сформированные, но содержащие отдельные пробелы представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Сформированы систематические представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	выполнение контрольных работ, лабораторная работа, тестовый контроль	7 семестр – зачет 8 семестр-экзамен
ПК-1.3. Осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	Не умеет и не готов осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	В целом успешное, но с систематическими ошибками осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	В целом успешное, но содержащее отдельные пробелы умение осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	Сформированное умение осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	выполнение контрольных работ, лабораторная работа, тестовый контроль	7 семестр – зачет 8 семестр-экзамен
ПК-1.4. Применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Не умеет применять прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Не достаточно хорошо применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	В целом успешно, но с отдельными пробелами применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Достаточно успешно применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	выполнение лабораторных работ ,выполнение контрольной работы, тестовый контроль	7 семестр – зачет 8 семестр-экзамен

ПК-2 Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности

Индикаторы достижения компетенции	Критерии оценивания результатов обучения				Средства оценивания результатов обучения	
	неудовлетв	удовлетв	хорошо	отлично	Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6	7
ПК-2.1. Разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Не знает основ управления проектами, и не реализует их с использованием современного программного обеспечения	Не разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Сформировано, но с отдельными пробелами разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Успешно разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	выполнение контрольных работ, лабораторная работа, тестовый контроль	7 семестр–зачет 8 семестр–экзамен
ПК-2.4. Оценивает временную и емкостную сложность программного обеспечения	Не умеет оценивать временную и емкостную сложность программного обеспечения	В целом успешно, но с систематическими ошибками оценивает временную и емкостную сложность программного обеспечения	В целом успешное, но содержащее отдельные пробелы умение оценивать временную и емкостную сложность программного обеспечения	В достаточной форме сформированное умение оценивать временную и емкостную сложность программного обеспечения	выполнение контрольных работ, лабораторная работа, тестовый контроль	7 семестр–зачет 8 семестр–экзамен
ПК-2.6. Работает с современными системами программирования и проектирования	Не умеет работать с современными системами программирования и проектирования	Недостаточно хорошо работает с современными системами программирования и проектирования	В целом успешно, но с отдельными пробелами работает с современными системами программирования и проектирования	Достаточно успешно работает с современными системами программирования и проектирования	выполнение лабораторных работ, выполнение индивидуальных заданий, тестирование	7 семестр–зачет 8 семестр–экзамен

1. Комплект контрольно-оценочных средств по дисциплине «Анализ данных и машинное обучение»

Вопросы к зачету по дисциплине Анализ данных и машинное обучение

2. Понятия «большие данные» (Big Data).
3. Инструменты работы с большими данными
4. Машинное обучение как область искусственного интеллекта.
5. Процесс построения модели ML
6. Ключевые характеристики больших данных (5 V: Volume, Velocity, Variety, Veracity, Value).
7. Назвать основные инструменты и технологии для работы с большими данными (например, Hadoop, Spark, posql-базы).
8. Python– интерпретируемый язык программирования. Синтаксис. Интерактивный режим, интерпретатор. Типы данных: целые числа, числа с плавающей точкой
9. Переменные Python.
10. Python– интерпретируемый язык программирования. Синтаксис. Интерактивный режим, интерпретатор. Типы данных: строки и списки
11. Python– интерпретируемый язык программирования. Синтаксис. Интерактивный режим, интерпретатор. Типы данных: множества, словари, кортежи
12. Сложная форма ветвления if– else. Инструкция if-elif-else.
13. Цикл for-цикл обхода, заданного множества элементов
14. Создание функций в Python. Локальные и глобальные переменные
15. Задачи машинного обучения. Обучением с учителем и без учителя
16. Задачи машинного обучения. Задача классификации
17. Задачи машинного обучения. Задача регрессии
18. Метрики эффективности для классификации
19. Метрики эффективности для регрессии
20. Задачи машинного обучения. Задача кластеризация
21. Меры расстояний в задачах кластеризации
22. Виды и алгоритмы кластеризация
23. Задачи машинного обучения. Задача понижения размерности
24. Алгоритмы отбора признаков в задачах понижения размерности
25. Метод фильтрации признаков в задачах понижения размерности.
26. Оберточные методы отбора признаков в задачах понижения размерности.
27. Встроенные методы отбора в задачах понижения размерности.
28. Алгоритмы выделения новых признаков в задачах понижения размерности.
29. Задачи машинного обучения. Задача определения тональности текстовых данных
30. Понятия параметров и аргументов функции.
31. Объяснить механизм передачи аргументов в функции (по значению/ссылке).
32. Продемонстрировать использование встроенных функций Python (например, len(), sum(), sorted()) на примерах.
33. Привести пример постановки задачи классификации/регрессии на реальных данных.
34. Объяснить понятие переобучения и методы борьбы с ним (регуляризация, кросс-валидация).
35. Этапы предварительной обработки данных (очистка, нормализация, кодирование категориальных переменных).
36. Привести примеры обработки пропущенных значений и выбросов.
37. Описать методы нормализации и стандартизации признаков.
38. Объяснить проблему высокой размерности данных и её последствия.
39. Дать определение понятию понижения размерности (dimensionality reduction).
40. Разграничить отбор признаков (feature selection) и выделение признаков (feature extraction).

41. Перечислить методы отбора признаков.
42. Описать алгоритм PCA (Principal Component Analysis) и его принцип работы.
43. Привести пример применения PCA для визуализации данных.
44. Назвать другие методы выделения признаков (t-SNE, Autoencoders).
45. Объяснить, как понижение размерности влияет на качество модели и скорость обучения.
46. Привести пример задачи, где понижение размерности критично для успеха модели.

Вопросы к экзамену
по дисциплине Анализ данных и машинное обучение

1. Понятие «ассоциативные правила»
2. Меры для определения ассоциаций
3. Формальное определение (ассоциативное правило)
4. Алгоритмы ассоциативных правил
5. Рекомендательные системы
6. Обучение с подкреплением Основные компоненты задачи обучения с подкреплением
7. Стратегии действий задачи обучения с подкреплением
8. Алгоритм обучения с подкреплением Q-обучение
9. Стратегия действий агента при Q-обучении
10. Оптимизация политики и Q-learning.
11. Policy gradients.
12. Подходы RL, основанные на моделях
13. Марковский процесс принятия решений
14. Алгоритмы машинного обучения. Решающие деревья
15. Алгоритмы машинного обучения. Метод опорных векторов
16. Алгоритмы машинного обучения. Наивный байесовский алгоритм
17. Алгоритмы машинного обучения. Кросс-валидация
18. Ансамбли моделей
19. Ансамбли моделей. Бэггинг
20. Ансамбли моделей. Бустинг
21. Ансамбли моделей. Стекинг
22. Основные типы рекомендательных систем
23. Принцип работы алгоритма Q-learning, включая формулу обновления Q-значений. Пример реальной задачи, решаемой методами обучения с подкреплением.
24. Основные структуры данных в Pandas и сравнить их с аналогами в реляционных БД.
25. Синтаксис фильтрации строк DataFrame по условию (например, возраст > 40).
26. Создание гистограммы распределения значений столбца с помощью Pandas/Matplotlib.
27. Описать шаги сохранения очищенного DataFrame в новый CSV-файл.
28. Основные методы исследования данных в Pandas и их вывод.
29. Способы обработки пропущенных значений в Pandas (isna(), dropna(), fillna()).
30. Группировка данных по столбцу и вычисление агрегатных функций (groupby() + sum(), mean()).
31. Библиотеки визуализации, интегрируемые с Pandas, и их ключевые преимущества.
32. Определение массива ndarray и его основные атрибуты (shape, dtype, ndim).
33. Создание массивов NumPy из списков, а также с помощью функций zeros(), ones(), arange().
34. Синтаксис индексации и срезов для одномерных и многомерных массивов.
35. Принцип векторизованных операций в NumPy и их преимущество перед циклами Python.
36. Использование универсальных функций (ufunc) — например, np.sin(), np.exp().
37. Методы изменения формы массива (reshape(), flatten(), transpose()).
38. Синтаксис объединения массивов (vstack(), hstack()) и их разделения (split()).
39. Агрегатные функции NumPy (sum(), mean(), std()) и параметр axis.
40. Способы приведения типов данных в массивах (astype()) и их влияние на потребление

памяти.

41. Привести пример вычисления накопленной суммы (`cumsum()`) и индексов экстремумов (`argmax()`).
42. Этапы предобработки текста для NLP (очистка, токенизация, удаление стоп-слов).
43. Суть стемминга и лемматизации, привести примеры их применения.
44. Модель Bag of Words (bow) и указать её ограничения.
45. Принцип расчёта TF-IDF и его функции
46. Векторизации текста с помощью TF-IDF с использованием `sklearn`.
47. Метрики качества для текстовой классификации (точность, полнота, F1-мера).
48. Архитектура простого классификатора текстов (например, Naive Bayes или Logistic Regression).
49. Задачи NLP, где критически важно учитывать порядок слов (и почему bow здесь не подходит).
50. Объяснить, как эмбединги слов (word embeddings) улучшают представление текста по сравнению с bow/TF-IDF.
51. Популярные предобученные модели для векторизации текста (Word2Vec, glove, BERT).
52. Структура искусственного нейрона (входы, веса, функция активации, выход).
53. Распространённые функции активации (sigmoid, relu, softmax) и их области применения.
54. Архитектура полносвязной нейронной сети (количество слоёв, связи между нейронами).
55. Определение задачи классификации и регрессии в контексте нейронных сетей.
56. Процесс прямого распространения сигнала (forward propagation) в сети.
57. Суть обратного распространения ошибки (backpropagation) и роль градиентного спуска.
58. Гиперпараметры обучения нейронной сети (скорость обучения, размер батча, количество эпох).
59. Сравнительная характеристика сверточных (CNN) и рекуррентных (RNN) нейронных сетей: области применения и ключевые отличия.
60. Основные фреймворки для построения нейронных сетей (Keras, tensorflow, pytorch) и их особенности.
61. Архитектура CNN для классификации изображений (схематично: входные слои, сверточные слои, пулинг, полносвязные слои).

Задачи к экзамену

по дисциплине Анализ данных и машинное обучение

Классификация с подбором гиперпараметров

Загрузите набор данных `iris` из `sklearn.datasets`.

Разделите данные на обучающую и тестовую выборки (70/30).

Обучите модель логистической регрессии и метод опорных векторов (SVM).

Для каждой модели подберите оптимальные гиперпараметры через `GridSearchCV`.

Сравните метрики: точность, полноту, F1-меру.

Сделайте вывод: какая модель лучше подходит для этой задачи?

Регрессия и оценка качества

Используйте датасет `boston` (или `california_housing`).

Постройте модель линейной регрессии и случайный лес для прогнозирования цены дома.

Оцените качество моделей через MSE, RMSE и R^2 .

Постройте график: истинные значения vs предсказания.

Объясните, почему одна модель может превосходить другую.

Работа с несбалансированными данными

Возьмите датасет с дисбалансом классов (например, creditcard fraud из Kaggle).

Вычислите долю положительного класса.

Обучите логистическую регрессию без учёта дисбаланса.

Примените методы балансировки (oversampling через SMOTE или class_weight).

Сравните F1-меру и ROC-AUC до и после балансировки.

Анализ данных о пассажирах Титаника

Загрузите CSV-файл titanic.csv.

Выведите: количество строк/столбцов, типы данных, число пропусков.

Отфильтруйте пассажиров старше 40 лет и женского пола.

Постройте гистограмму возраста выживших и погибших.

Сохраните очищенные данные в новый файл titanic_clean.csv.

Манипуляции с массивами NumPy

Создайте массив 6×6 со случайными числами от 0 до 1.

Замените значения > 0.7 на 1, остальные — на 0.

Найдите сумму по строкам и столбцам.

Трансформируйте массив в 1D, отсортируйте по убыванию.

Вычислите медиану и 75-й перцентиль полученного вектора.

Группировка и агрегация в Pandas

Загрузите данные о продажах (столбцы: date, product, category, revenue).

Преобразуйте date в datetime.

Сгруппируйте по category и посчитайте: общую выручку, среднее значение, количество транзакций.

Выведите топ-3 категории по выручке.

Постройте столбчатую диаграмму выручки по категориям.

Векторизация текста

Возьмите 10–15 коротких текстов (например, отзывы о товарах).

Проведите предобработку: приведение к нижнему регистру, удаление пунктуации, стемминг.

Векторизуйте тексты через Bag of Words и TF-IDF (используйте sklearn).

Сравните разреженность матриц (доля нулевых элементов).

Объясните, в чём преимущество TF-IDF перед BoW.

Классификация текстов

Используйте датасет 20 newsgroups (темы: спорт, политика, технологии).

Разделите на обучающую/тестовую выборки.

Обучите классификатор (например, Naïve Bayes) на TF-IDF-представлении.

Выведите матрицу ошибок и F1-меру по классам.

Назовите 2 способа улучшить качество модели.

Построение полносвязной нейронной сети

Используйте датасет mnist (цифры 0–9).

Нормализуйте пиксели изображений (деление на 255).

Создайте модель с 2 скрытыми слоями (128 и 64 нейрона, активация ReLU).

Обучите модель 10 эпох, используйте оптимизатор Adam и категориальную кросс-энтропию.

Оцените точность на тестовой выборке.

Постройте график изменения потерь и точности по эпохам.

Обучение с подкреплением (Q-learning)

Реализуйте среду 4×4 с целью в правом верхнем углу.

Действия агента: вверх, вниз, влево, вправо (штраф -1 за шаг, награда +10 за цель).

Обучите агента через Q-learning ($\alpha = 0.1$, $\gamma = 0.9$, $\epsilon = 0.2$).

Постройте график: номер эпизода vs суммарная награда.

Проанализируйте, как меняется политика агента после 100, 500 и 1000 эпизодов.

Рекомендательная система на основе коллаборативной фильтрации

Используйте датасет ml-100k (рейтинги фильмов).

Разделите на обучающую/тестовую выборки.

Постройте модель user-based kNN ($k = 20$).

Оцените RMSE на тестовых данных.

Выведите 5 рекомендаций для пользователя с ID = 10.

Ассоциативные правила в ритейле

Загрузите транзакционные данные (столбцы: transaction_id, item).

Примените алгоритм Apriori ($\min_support = 0.02$, $\min_confidence = 0.4$).

Найдите 5 правил с максимальным lift.

Интерпретируйте одно правило: какие товары рекомендуется размещать рядом?

Анализ временных рядов

Загрузите данные о ежедневных продажах товара за год.

Визуализируйте временной ряд.

Вычислите скользящее среднее за 7 дней.

Выделите сезонность (например, недельную).

Предскажите продажи на следующие 7 дней (простой метод: среднее за последние 14 дней).

СЕВЕРО-КАВКАЗСКАЯ ГОСУДАРСТВЕННАЯ АКАДЕМИЯ

Кафедра Цифровой инженерии и сетевых технологий

20 - 20 учебный год

Экзаменационный билет № 1

по дисциплине Анализ данных и машинное обучение
для обучающихся направления подготовки 09.03.04 Программная инженерия

1. Алгоритмы машинного обучения. Кросс-валидация
2. Основные методы исследования данных в Pandas и их вывод.
3. Классификация текстов
Используйте датасет 20 newsgroups (темы: спорт, политика, технологии).
Разделите на обучающую/тестовую выборки.
Обучите классификатор (например, Naive Bayes) на TF-IDF-представлении.
Выведите матрицу ошибок и F1-меру по классам.
Назовите 2 способа улучшить качество модели.

Зав. кафедрой

Эльканова Л.М.

Контрольные работы

Работа с файлом данных «Титаника»

Приведенные ниже задания основаны на данных 'titanic.csv', где содержатся сведения о пассажирах «Титаника».

Задачи:

1. Какое количество мужчин и женщин плыло на корабле?
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров. Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров? Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров. В качестве ответа приведите два числа через пробел.
5. Коррелируют ли число братьев/сестер/супругов с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch.

Тестовые вопросы
по дисциплине: «Анализ данных и машинное обучение»

7 семестр

1. Какой инструмент традиционно НЕ относится к экосистеме для работы с большими данными (Big Data)?
 - a) Apache Hadoop
 - b) Microsoft Excel
 - c) Apache Spark
 - d) NoSQL-базы данных (Cassandra, MongoDB)
2. Допишите правильный ответ. Область информатики, связанная с исследованием и разработкой технологий для сбора, хранения, обработки и анализа очень больших объемов данных – это... .. (на англ.)
3. Машинное обучение (ML) является подразделом:
 - a) Системного программирования
 - b) Искусственного интеллекта (ИИ)
 - c) Компьютерной графики
 - d) Сетевых технологий
4. Установите соответствие между основными характеристиками больших данных, известными как 4v и их описанием.
 - a) Объем (Volume) — огромные объемы данных, которые могут измеряться в терабайтах, петабайтах и даже эксабайтах.
 - b) Скорость (Velocity) — высокая скорость генерации и обработки данных, что требует мгновенной реакции.
 - в) Разнообразие (Variety) — данные могут быть представлены в различных форматах: текст, изображения, видео, аудио, числовые данные и т.д.
 - г) Достоверность (Veracity) — качество данных, их точность и надежность, что важно для анализа и принятия решений.
5. Какой этап является первым в типичном процессе построения модели машинного обучения?
 - a) Выбор алгоритма
 - b) Сбор и подготовка данных
 - c) Развертывание модели
 - d) Оценка качества модели
6. Допишите правильный ответ Ключевое отличие машинного обучения от классического программирования - способность ...: обычная программа работает по заданным правилам и не меняется, ML-модель улучшается по мере накопления данных и опыта
7. Допишите правильный ответ. ... -подмножество элементов, отобранных из более крупной совокупности (генеральной совокупности) для проведения исследования.
8. Какой тип данных в Python является неизменяемым (immutable)?
 - a) Список (list)
 - b) Кортеж (tuple)
 - c) Словарь (dict)
 - d) Множество (set)
9. Какой результат будет у выражения `type(5 / 2)` в Python 3.x?
 - a) `<class 'int'>`
 - b) `<class 'float'>`
 - c) `<class 'str'>`
 - d) Возникнет ошибка
10. Какой метод используется для добавления элемента в конец списка?
 - a) `.insert()`
 - b) `.append()`
 - c) `.add()`
 - d) `.push()`

11. Чем принципиально отличается кортеж (tuple) от списка (list)?
 - a) Кортеж может хранить только числа
 - b) Кортеж является неизменяемым (immutable)
 - c) Список занимает меньше памяти
 - d) У кортежа нет индексов
12. Какая конструкция используется для итерации по последовательности в Python?
 - a) while
 - b) for
 - c) loop
 - d) iterate
13. Как объявить функцию в Python?
 - a) function my_func():
 - b) def my_func():
 - c) declare my_func():
 - d) func my_func():
14. Что такое "локальная переменная" функции?
 - a) Переменная, объявленная с ключевым словом global
 - b) Переменная, объявленная внутри функции и не видимая за ее пределами
 - c) Переменная, которая всегда равна None
 - d) Системная переменная интерпретатора
15. Выберите строку, которая импортирует пакет для визуализации данных в Python из нужного модуля
 - a) import numpy as np
 - b) import matplotlib.pyplot as plt
 - c) import pandas as pd
 - d) from sklearn.datasets import load_breast_cancer
16. Как называется задача, где целевая переменная является категориальной (например, "спам"/"не спам")?
 - a) Регрессия
 - b) Кластеризация
 - c) Классификация
 - d) Понижение размерности
17. Допишите правильный ответ .В матрице ошибок классификации число объектов, которые алгоритм корректно отнес к классу - это
18. Выберите правильный синтаксис импорта нужной функции для разделения данных на обучающую и тестовую
 - a) from sklearn.model_selection import train_test_split
 - б) sklearn.model_selection import train_test_split
 - в) from sklearn.model_selection import train_test_
 - г) from sklearn.selection import train_test_split
19. Допишите правильный ответ .В матрице ошибок классификации Число объектов, которые алгоритм ошибочно не отнес к классу, хотя они должны были быть к нему отнесены – это
20. Какая метрика НЕ используется для оценки задач регрессии?
 - a) Среднеквадратичная ошибка (MSE)
 - b) Средняя абсолютная ошибка (MAE)
 - c) Точность (Accuracy)
 - d) Коэффициент детерминации (R^2)
21. Алгоритм k-NN (k ближайших соседей) является примером:
 - a) Ансамблевого метода
 - b) Метода, основанного на экземплярах (instance-based)
 - c) Градиентного бустинга
 - d) Нейронной сети

22. Как называется задача разбиения набора данных на группы схожих объектов?
- Регрессия
 - Классификация
 - Кластеризация
 - Ассоциация
23. Допишите правильный ответ. Расстояние L_∞ является одной из мер расстояния, которая измеряет разницу между соответствующими координатами двух точек в пространстве
24. По степени вложенности кластеров выделяются следующие виды алгоритмов кластеризации:
- Плоские (flat) — разбивают данные на один уровень кластеров, не учитывая их иерархию
 - Иерархические (hierarchical) — разбивают данные на несколько уровней кластеров, учитывая их иерарх
 - Перекрывающиеся (overlapping) — каждый объект может принадлежать нескольким кластерам (например, MCOKE).
 - Исключающие (exclusive) — каждый объект принадлежит только одному кластеру
25. Вы построили модель логистической регрессии. Какая метрика будет наиболее информативной для оценки её качества на сбалансированном датасете?
- R^2
 - MSE
 - Accuracy (Точность) или F1-score
 - Support (Поддержка)
26. В чем основное различие между задачами классификации и кластеризации?
- В классификации есть признаки, в кластеризации — нет
 - В классификации известны метки классов для обучения, в кластеризации — нет
 - Кластеризация использует только линейные модели
 - Классификация всегда дает вероятности на выходе
27. Допишите правильный ответ. Расстояние городских кварталов (расстояние L1) или, измеряет дистанцию между двумя точками в пространстве, учитывая только перемещение по осям координат
28. Допишите правильный ответ. ...- древовидная диаграмма, которая показывает иерархическую связь между объектами.
29. Допишите правильный ответ. Расстояние L2, известное как евклидово расстояние - наиболее распространенная мера расстояния в евклидовой геометрии, измеряет кратчайшее расстояние между двумя
30. Вы используете K-Means. Как выбирается параметр K (число кластеров)?
- Он всегда равен 3
 - С помощью метода "локтя" (elbow method) по инерции (within-cluster sum of squares)
 - Случайным образом
 - По максимальной схожести всех точек
31. Что такое "переобучение" (overfitting) модели?
- Модель слишком проста и плохо работает даже на обучающих данных
 - Модель слишком сложна, идеально подстраивается под обучающие данные, но плохо обобщает на новые
 - Модель не сходится в процессе обучения
 - Модель использует слишком мало признаков
32. Основная структура данных в Pandas для работы с таблицами — это:
- Series
 - DataFrame
 - ndarray
 - List

33. Какой метод Pandas используется для чтения CSV-файла?
- pd.open_csv()
 - pd.read_excel()
 - pd.read_csv()
 - pd.from_csv()
34. Что делает метод df.info()?
- Показывает краткую сводку о DataFrame: типы столбцов, количество ненулевых значений
 - Выводит статистику по числовым столбцам
 - Показывает первые 5 строк
 - Создает гистограммы для всех столбцов
35. Какая библиотека Python является основой для большинства инструментов визуализации?
- Seaborn
 - Plotly
 - Matplotlib
 - Bokeh
36. Библиотека Seaborn:
- Работает независимо от Matplotlib
 - Предоставляет более высокоуровневый и статистически ориентированный интерфейс для построения графиков
 - Используется только для 3D-графиков
 - Не поддерживает работу с Pandas DataFrame
37. Основным объектом для хранения данных в NumPy — это:
- List
 - ndarray (N-dimensional array)
 - Series
 - Tensor
38. Допишите правильный ответ. Для оценки уже построенных моделей используется не функция ошибки, а так называемые - специальные функции, которые показывают, насколько эффективна уже готовая, обученная модель.
39. Допишите правильный ответ. Метод библиотеки Pandas в Python, который позволяет просмотреть первые несколько строк DataFrame или Series это ...
40. Допишите правильный ответ. Эффективность - это свойство модели машинного обучения давать предсказания значения... .., как можно ближе к реальным данным.
41. Кластеризация – это
- Задача обучения с учителем, где нужно предсказать категориальную переменную.
 - Задача обучения без учителя, где нужно разбить данные на группы схожих объектов.
 - Задача обучения с учителем, где нужно предсказать непрерывную переменную.
 - Задача обучения без учителя, где нужно предсказать пропущенные значения.
42. В методе K-means для определения схожести объектов обычно используется:
- Косинусная мера
 - Евклидово расстояние
 - Манхэттенское расстояние
 - Все вышеперечисленное могут использоваться, но чаще всего евклидово.
43. Что такое "центроид" в алгоритме K-means?
- Среднее арифметическое всех точек в кластере.
 - Медиана всех точек в кластере.
 - Точка, наиболее удаленная от всех остальных в кластере.
 - Случайно выбранная точка в кластере.
44. Алгоритм DBSCAN группирует точки на основе:
- Заданного числа кластеров.
 - Плотности точек в пространстве.
 - Иерархии слияния или разделения кластеров.
 - Вероятности принадлежности к кластеру.

45. Какой из перечисленных алгоритмов НЕ является алгоритмом классификации?
- Линейная регрессия
 - Логистическая регрессия
 - Решающие деревья
 - Метод опорных векторов (SVM)
46. Допишите правильный ответ. ... - это метрика, используемая для оценки качества классификации, особенно в задачах с несбалансированными классами. Она объединяет в себе две другие важные метрики - точность (Precision) и полноту (Recall).
47. Какой алгоритм классификации основан на теореме Байеса и предположении о независимости признаков?
- Наивный байесовский классификатор
 - Метод k-ближайших соседей
 - Логистическая регрессия
 - Нейронные сети
48. Что такое confusion matrix (матрица ошибок)?
- Матрица, показывающая распределение правильных и неправильных прогнозов модели по классам.
 - Матрица, отображающая корреляции между признаками.
 - Матрица, используемая для уменьшения размерности данных.
 - Матрица, используемая для кластеризации данных.
49. Допишите правильный ответ.... - это график, который иллюстрирует производительность классификационной модели при всех возможных порогах классификации. Идеальная модель классификации будет стремиться к точке в верхнем левом углу графика, где TPR равно 1, а FPR равно 0.
50. Допишите правильный ответ. В библиотеке ... для предварительной обработки данных доступны методы, связанные с масштабированием, кодированием, заполнением пропусков и уменьшением размерности. Эти методы помогают преобразовать исходные данные в формат, подходящий для алгоритма обучения.
51. Что оценивает коэффициент детерминации R^2 ?
- Долю дисперсии зависимой переменной, объясненную моделью.
 - Среднюю абсолютную ошибку модели.
 - Среднеквадратичную ошибку модели.
 - Корреляцию между признаками.
52. Какая из метрик является метрикой ошибки в задачах регрессии? (выберите все подходящие)
- MAE (Mean Absolute Error)
 - MSE (Mean Squared Error)
 - RMSE (Root Mean Squared Error)
 - Accuracy
53. Установите последовательность этапов исследования
- Загрузка данных
 - Предварительная обработка данных
 - Исследовательский анализ данных
 - Отбор и выделение признаков
 - Обучение и оценка качества модели
54. Допишите правильный ответ. Среднеквадратическая ошибка (Mean Squared Error, MSE)- это метрика в статистике и машинном обучении, которая измеряет среднее значение ... между прогнозируемыми и фактическими значениями
55. Допишите правильный ответ Компонент target_names содержит названия ...
56. Допишите правильный ответ Если в наборе данных пропущен существенный процент значений, то мы не сможем построить корректную модель можно воспользоваться функцией ...
57. Что такое регуляризация в регрессии?
- Метод уменьшения сложности модели для предотвращения переобучения путем

- добавления штрафа за большие коэффициенты.
- b) Метод увеличения точности модели на обучающей выборке.
 - c) Метод отбора признаков на основе их важности.
 - d) Метод преобразования признаков для уменьшения выбросов.
58. Допишите правильный ответ. MAE (Mean Absolute Error) — метрика в машинном обучении, которая измеряет среднюю абсолютную ошибку между ... и ... значениями
59. Допишите правильный ответ. R^2 (коэффициент детерминации, R-квадрат) - показатель оценки эффективности моделей машинного обучения на основе регрессии. Он измеряет долю ... зависимой переменной, объясняемую рассматриваемой моделью
60. Какие из перечисленных методов являются методами регуляризации в регрессии? (выберите все подходящие)
- a) L1-регуляризация (Lasso)
 - b) L2-регуляризация (Ridge)
 - c) Elastic Net
 - d) Удаление признаков с высокой корреляцией

8 семестр

1. Основная цель анализа ассоциативных правил в данных о покупках:
 - a) Предсказать, кто купит товар
 - b) Найти товары, которые часто покупаются вместе
 - c) Классифицировать клиентов на группы
 - d) Спрогнозировать выручку от каждого товара
2. Правило {Молоко, Хлеб} → {Яйца} имеет Support = 0.1 и Confidence = 0.8. Что означает Confidence = 0.8?
 - a) 80% всех транзакций содержат все три товара
 - b) Из всех транзакций с молоком и хлебом 80% содержат яйца
 - c) 80% транзакций содержат либо молоко, либо хлеб, либо яйца
 - d) Товары встречаются вместе в 80% случаев
3. Какая из перечисленных метрик НЕ используется при анализе ассоциативных правил?
 - a) Support (Поддержка)
 - b) Confidence (Достоверность)
 - c) Lift (Лифт)
 - d) F1-score
4. Правило с высоким Support, но низким Lift, скорее всего:
 - a) Очень интересное и полезное
 - b) Тривиальное, так как товары и так часто покупаются по отдельности
 - c) Редкое, но значимое
 - d) Противоречивое
5. FP-Growth в сравнении с Apriori:
 - a) Требуется больше памяти
 - b) Обычно работает быстрее, так как не требует многократного сканирования базы
 - c) Всегда находит больше правил
 - d) Не использует метрики Support и Confidence
6. Концепция "сильного правила" (strong rule) обычно подразумевает, что:
 - a) Support и Confidence выше заданных порогов
 - b) Только Support выше порога
 - c) Только Confidence выше порога
 - d) Lift точно равен 1
7. Анализ ассоциативных правил можно применять для:
 - a) Составления рекомендаций "сопутствующие товары"
 - b) Планирования размещения товаров в магазине

- c) Прогнозирования спроса на отдельный товар
 - d) Определения целевых групп клиентов
8. Система рекомендаций, которая предлагает товары, похожие на те, что пользователь уже покупал, основана на:
- a) Коллаборативной фильтрации
 - b) Контентной фильтрации
 - c) Гибридном подходе
 - d) Популярности товаров
9. Коллаборативная фильтрация user-based работает по принципу:
- a) Найти пользователей, похожих на текущего, и рекомендовать то, что понравилось им
 - b) Найти товары, похожие на те, что нравятся пользователю
 - c) Использовать признаки товаров и профиль пользователя
 - d) Рекомендовать самые популярные товары
10. Матрица пользователь-предмет (user-item matrix) в коллаборативной фильтрации обычно является:
- a) Полностью заполненной
 - b) Разреженной (sparse)
 - c) Квадратной
 - d) Симметричной
11. SVD (Singular Value Decomposition) в рекомендательных системах используется для:
- a) Уменьшения размерности и выявления скрытых факторов
 - b) Кластеризации пользователей
 - c) Ускорения вычисления евклидова расстояния
 - d) Нормализации оценок
12. Метрика RMSE (Root Mean Square Error) в рекомендательных системах оценивает:
- a) Скорость работы алгоритма
 - b) Точность предсказания оценок
 - c) Разнообразие рекомендаций
 - d) Степень персонализации
13. Гибридные рекомендательные системы:
- a) Всегда работают медленнее, чем отдельные подходы
 - b) Комбинируют несколько методов для улучшения качества рекомендаций
 - c) Требуют меньше данных
 - d) Легче поддаются интерпретации
14. Какие из перечисленных метрик важны для оценки рекомендательных систем?
- a) Точность (precision@k, RMSE)
 - b) Полнота (recall@k)
 - c) Разнообразие (diversity) и новизна (novelty)
 - d) Все вышеперечисленное
15. Основная цель агента в обучении с подкреплением:
- a) Классифицировать состояния среды
 - b) Максимизировать накопленное вознаграждение (return)
 - c) Минимизировать ошибку предсказания
 - d) Найти оптимальное разделение данных
16. Марковский процесс принятия решений (MDP) характеризуется:
- a) Только текущим состоянием (свойство Маркова)
 - b) Набором состояний, действий, вероятностей переходов и вознаграждений
 - c) Необходимостью полного знания среды
 - d) Все вышеперечисленное
17. Алгоритм Q-learning является:
- a) Методом based на модели (model-based)
 - b) Методом без модели (model-free)
 - c) Методом on-policy
 - d) Методом только для непрерывных пространств действий

18. Дилемма "exploration vs exploitation" в RL — это:
- a) Выбор между изучением новых действий и использованием известных лучших
 - b) Выбор между скоростью и памятью
 - c) Выбор между точностью и полнотой
 - d) Выбор между обучением и тестированием
19. Политика (policy) в RL — это:
- a) Стратегия выбора действий агентом
 - b) Функция, оценивающая качество состояния
 - c) Модель переходов среды
 - d) Распределение вознаграждений
20. Алгоритмы Policy Gradient работают:
- a) Непосредственно оптимизируя параметры политики
 - b) Оценивая функцию ценности
 - c) Строя модель среды
 - d) Только с дискретными действиями
21. Глубокое обучение с подкреплением (Deep RL) — это:
- a) Использование нейронных сетей для аппроксимации функции ценности или политики
 - b) RL с большим количеством состояний
 - c) RL с непрерывным пространством действий
 - d) Комбинация RL и supervised learning
22. Какие из перечисленных являются реальными применениями RL?
- a) Обучение игровым AI (AlphaGo, Dota 2)
 - b) Управление роботами и беспилотными автомобилями
 - c) Оптимизация систем рекомендаций
 - d) Все вышеперечисленное
23. Какова основная цель использования PCA (метода главных компонент)?
- a) Увеличение числа признаков
 - b) Уменьшение размерности данных при сохранении максимальной дисперсии
 - c) Классификация объектов
 - d) Поиск ассоциативных правил
24. Как называется метод понижения размерности, который ищет новые ортогональные признаки (главные компоненты)?
- a) LDA (Linear Discriminant Analysis)
 - b) PCA (Principal Component Analysis)
 - c) t-SNE
 - d) ISOMAP
25. Какой тип рекомендательных систем использует матрицу "пользователь-объект" и ищет скрытые факторы?
- a) Контентная фильтрация
 - b) Коллаборативная фильтрация на основе памяти
 - c) Коллаборативная фильтрация на основе модели (Matrix Factorization)
 - d) Гибридные системы
26. В обучении с подкреплением агент получает награду (reward) за:
- a) Быстрое завершение эпизода
 - b) Выполнение действий, ведущих к цели
 - c) Использование всех возможных действий
 - d) Бездействие
27. Алгоритм Q-Learning относится к:
- a) Контролю по стратегии (Policy Gradient)
 - b) Контролю по ценности (Value-Based)
 - c) Актор-критическим методам (Actor-Critic)
 - d) Методам на основе моделей
28. Допишите . Обучение с подкреплением – это обучение путем взаимодействия со ...

29. Допишите.... - сущность, которая принимает решения или действия в среде при обучении с подкреплением
30. Допишите. Конкретное условие или конфигурация среды и агента в определенный момент времени – это ...
31. Допишите ... -это стратегия агента, которая определяет, какое действие следует предпринять в каждом состоянии.
32. Допишите .Марковский процесс принятия решений (Markov Decision Process, MDP) представляет собой, которая описывает основные составляющие задачи.
33. Допишите— метод машинного обучения, где несколько моделей обучаются для решения одной и той же проблемы и объединяются для получения лучших результатов.
34. Допишите ... - ансамбль моделей одного вида, обучающихся параллельно и независимо друг от друга, на различных случайных выборках одного и того же обучающего множества
35. Допишите . К методам принятия решений при бэггинге относятся ... ,простое большинство, взвешивание классификаторов.
36. Допишите. ... - это процедура последовательного направленного построения ансамбля моделей машинного обучения, когда каждый следующий алгоритм стремится компенсировать ошибки предыдущих алгоритмов.
37. Вычисление взвешенной суммы входов и смещения в искусственном нейроне называется:
- Функция активации
 - Линейная комбинация (или агрегация)
 - Операция свертки
 - Обратное распространение
38. Какая функция активации чаще всего используется на выходном слое для задачи бинарной классификации?
- ReLU
 - Sigmoid
 - Tanh
 - Sigmoid (или, строго говоря, логистическая функция)
39. Архитектура, где каждый нейрон слоя связан со всеми нейронами следующего слоя, называется:
- Сверточной нейронной сетью (CNN)
 - Полносвязной нейронной сетью (Fully Connected, Dense)
 - Рекуррентной нейронной сетью (RNN)
 - Автоэнкодером
40. Какой фреймворк предоставляет высокоуровневый API для построения нейронных сетей и часто работает поверх TensorFlow?
- PyTorch
 - Keras
 - Scikit-learn
 - Caffe
41. Сверточные нейронные сети (CNN) особенно эффективны для задач:
- Обработки временных рядов
 - Компьютерного зрения (анализ изображений)
 - Кластеризации текстов
 - Обучения с подкреплением
42. Рекуррентные нейронные сети (RNN) предназначены для работы с:
- Изображениями
 - Последовательностями данных (текст, временные ряды)
 - Табличными данными
 - Графами
43. Прямое распространение (Forward propagation) — это процесс:
- Обновления весов сети
 - Вычисления выхода сети на основе входных данных и текущих весов

- c) Оценки градиента функции потерь
 - d) Регуляризации модели
44. Алгоритм обратного распространения ошибки (Backpropagation) используется для:
- a) Классификации новых данных
 - b) Вычисления градиентов функции потерь по весам сети
 - c) Инициализации весов сети
 - d) Визуализации работы сети
45. При стандартизации признаков с помощью StandardScaler из sklearn, преобразование основывается на:
- a) Минимуме и максимуме признака
 - b) Среднем значении и стандартном отклонении признака
 - c) Медиане и межквартильном размахе
 - d) Просто на центрировании
46. При построении гистограммы для визуализации распределения непрерывного признака, чаще всего, используется:
- a) plt.scatter()
 - b) plt.pie()
 - c) plt.hist() или sns.histplot()
 - d) plt.boxplot()
47. Какой объект Pandas позволяет выполнять группировку данных и агрегацию (например, вычисление среднего по группам)?
- a) df.sort_values()
 - b) df.groupby()
 - c) df.pivot_table()
 - d) df.merge()
48. Для чего используется train_test_split из sklearn.model_selection?
- a) Для разделения данных на обучающую и тестовую выборки случайным образом
 - b) Для разбиения данных на K фолдов для кросс-валидации
 - c) Для центрирования \
 - a. Для нормализации данных
49. Допишите. Объединение разнородных алгоритмов для достижения максимальной точности - это ...
50. Допишите представляют собой механизм нахождения логических закономерностей между связанными элементами (событиями или объектами)
51. Допишите ... как мера для определения ассоциаций показывает то, как часто данный товарный набор появляется, что измеряется долей покупок, в которых он присутствует.
52. Допишите ... как мера для определения ассоциаций показывает, как часто товар Y появляется вместе с товаром X, что выражается как $\{X \rightarrow Y\}$. Это измеряется долей их одновременных появлений.
53. Допишите ... как мера для определения ассоциаций отражает то, как часто товары X и Y появляются вместе, одновременно учитывая, с какой частотой появляется каждый из них.
54. Допишите Задача построения различных предложений клиенту на основе ассоциативных правил – это задача построения... ..
55. Допишите — это алгоритм, который делает предсказания на основе серии вопросов об объекте.
56. Допишите Конечные (или, терминальные) узлы решающего дерева называют ..., каждый из которых представляет конечный результат последовательного принятия решений
57. Допишите ... — это параметр модели машинного обучения, значение которого устанавливается перед началом процесса обучения и не изменяется алгоритмом. В отличие от обычных параметров, вычисляемых в процессе обучения, они определяют структуру модели и способ её обучения
58. Допишите Метод опорных векторов — набор схожих алгоритмов обучения с учителем, использующихся для задач ... и регрессионного анализа.

59. Допишите ...-...— это метод, предназначенный для оценки качества работы модели, широко применяемый в машинном обучении.

60. Допишите Leave-one-out , Stratified K-folds , Nested K-folds , K-folds , Stratified K-folds -это популярные методы ...-...

1. Методические материалы, определяющие процедуры оценивания компетенции

1.1 Критерии оценивания качества выполнения лабораторных работ

Оценка «зачтено» выставляется обучающемуся, если лабораторная работа выполнена правильно и студент ответил на все вопросы, поставленные преподавателем на защите.

Оценка «не зачтено» выставляется обучающемуся, если лабораторная работа выполнена не правильно или студент не проявил глубоких теоретических знаний при защите работы.

1.2 Критерии оценивания тестирования

При тестировании все верные ответы берутся за 100%.

90%-100% отлично

75%-90% хорошо

60%-75% удовлетворительно

менее 60% неудовлетворительно

1.3 Критерии оценивания контрольных работ

Оценка «отлично» выставляется за глубокое знание предусмотренного программой материала, за умение четко, лаконично и логически последовательно отвечать на поставленные вопросы.

Оценка «хорошо» – за твердое знание основного (программного) материала, за грамотные, без существенных неточностей ответы на поставленные вопросы.

Оценка «удовлетворительно» – за общее знание только основного материала, за ответы, содержащие неточности или слабо аргументированные, с нарушением последовательности изложения материала.

Оценка «неудовлетворительно» – за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в материале, за незнание основных понятий дисциплины.

1.4 Критерии оценивания ответа на зачете

Оценка «зачтено» выставляется обучающемуся за общее знание основного материала, включая расчеты (при необходимости), за грамотные, без существенных неточностей ответы на поставленные вопросы, за умение применять теоретические положения для решения практических задач.

Оценка «не зачтено» выставляется обучающемуся за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в расчетах, за незнание основных понятий дисциплины.

1.5 Критерии оценивания результатов освоения дисциплины на экзамен

Оценка «отлично» выставляется за глубокое знание предусмотренного программой материала, содержащегося в основных и дополнительных рекомендованных литературных источниках, за умение четко, лаконично и логически последовательно отвечать на поставленные вопросы, за умение анализировать изучаемые явления в их взаимосвязи и диалектическом развитии, применять теоретические положения при решении практических задач.

Оценка «хорошо» – за твердое знание основного (программного) материала, включая расчеты (при необходимости), за грамотные, без существенных неточностей ответы на поставленные вопросы, за умение применять теоретические положения для

решения практических задач.

Оценка **«удовлетворительно»** – за общее знание только основного материала, за ответы, содержащие неточности или слабо аргументированные, с нарушением последовательности изложения материала, за слабое применение теоретических положений при решении практических задач.

Оценка **«неудовлетворительно»** – за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в расчетах, за незнание основных понятий дисциплины.