

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Нижегородский государственный технический университет
им. Р.Е. Алексеева» (НГТУ)

Передовая инженерная школа атомного машиностроения и систем
высокой плотности энергии

УТВЕРЖДАЮ:

Директор ПИШ:

_____ Тумасов А.В.
подпись ФИО

“ 21 ” _____ 06 _____ 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.Б.6 Анализ больших данных

(индекс и наименование дисциплины по учебному плану)

для подготовки магистров

Направление подготовки: 22.04.01 Материаловедение и технологии материалов

Направленность: Материалы для высокотемпературных ядерных реакторов

Форма обучения: очная

Год начала подготовки 2024

Выпускающая кафедра «Материаловедение, технологии материалов и термическая обработка металлов» (МТМиТОМ)

Кафедра-разработчик ИСУ

Объем дисциплины 108/3
часов/з.е

Промежуточная аттестация зачёт с оценкой

Разработчик: Багиров М.Б., старший преподаватель

Нижний Новгород, 2024

Рабочая программа дисциплины: разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО 3++) по направлению подготовки 22.04.01 Материаловедение и технологии материалов, утвержденного приказом МИНОБРНАУКИ РОССИИ от 24 апреля 2018 года № 306 на основании учебного плана принятого УМС НГТУ, протокол от 23.04.24 №14

Рабочая программа одобрена на заседании кафедры разработчика протокол от 15.05.2024 №9

Зав. кафедрой к.т.н, доцент Тимофеева О.П. _____
(подпись)

Программа рекомендована к утверждению ученым советом института ИФХТиМ, протокол от 20.06.2024 №4

Председатель УМС, директор ИФХТиМ Мацулевич Ж.В. _____
(подпись)

Рабочая программа зарегистрирована в УМУ, регистрационный № 22.04.01-п-5
Начальник МО _____ Н.Р. Булгакова

Заведующая отделом комплектования НТБ _____ Н.И. Кабанина
(подпись)

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	4
1.1 Цель освоения дисциплины.....	4
1.2 Задачи освоения дисциплины (модуля)	4
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	4
3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	5
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	7
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ ПО СЕМЕСТРАМ	7
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ, СТРУКТУРИРОВАННОЕ ПО ТЕМАМ.....	8
5. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	13
5.1 ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ	13
5.2 ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ	14
6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	16
7. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	17
7.1 ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ	17
7.2 ПЕРЕЧЕНЬ СВОБОДНО РАСПРОСТРАНЯЕМОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ	17
7.3 ПЕРЕЧЕНЬ СОВРЕМЕННЫХ ПРОФЕССИОНАЛЬНЫХ БАЗ ДАННЫХ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ.....	17
8. ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОВЗ.....	18
9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ, НЕОБХОДИМОЕ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....	18
10. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ	20
10.1 ОБЩИЕ МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ, ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....	20
10.2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ЗАНЯТИЙ ЛЕКЦИОННОГО ТИПА.....	21
10.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ НА ЛАБОРАТОРНЫХ РАБОТАХ.....	21
10.4 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ НА ПРАКТИЧЕСКИХ ЗАНЯТИЯХ	21
10.5 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ НА КУРСОВОЙ РАБОТЕ	21
10.6 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО САМОСТОЯТЕЛЬНОЙ РАБОТЕ ОБУЧАЮЩИХСЯ	21
11. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ	22
11.1 ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА В ХОДЕ КОНТРОЛЯ ТЕКУЩЕЙ УСПЕВАЕМОСТИ.....	22
11.2 ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА В ХОДЕ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ.....	22

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1 Цель освоения дисциплины

Целями изучения дисциплины являются изучение методов и подходов к анализу данных различного объема, включая предварительную обработку данных и статистический анализ, освоение различных моделей машинного обучения, предназначенных для решения задач кластеризации, классификации и регрессии и применение их для решения прикладных задач из различных сфер человеческой деятельности.

1.2 Задачи освоения дисциплины (модуля)

Дисциплина «Анализ больших данных» способствует подготовке студентов к решению следующих профессиональных задач:

1. Разработка программного кода для эффективной обработки распределенных данных большого объема в сочетании с распределенным хранением данных в Hadoop кластере.
2. Построение и применение на практике описательных и прогнозных моделей интеллектуального анализа больших данных и машинного обучения.
3. Использование программных средств визуализации и интерактивного исследования больших данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Учебная дисциплина «Анализ больших данных» Б1.Б.6 включена в обязательный перечень дисциплин.

Дисциплина базируется на следующих дисциплинах в объеме программы магистратуры. Предшествующими курсами, на которых непосредственно базируется дисциплина «Анализ больших данных», являются бакалаврские курсы, в частности: «Программирование», «Математический анализ», «Дискретная математика» и другие дисциплины.

Дисциплина «Анализ больших данных» является основополагающей для преддипломной практики и выполнения выпускной квалификационной работы.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Дисциплина «Анализ больших данных» формирует компетенции УК-1 совместно с дисциплинами и практиками, указанными в таблице 3.1.

Дисциплинарная часть компетенции УК-1 «Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий»: способен понимать и применять на практике методы интеллектуального анализа и обработки больших данных.

Таблица 3.1 – Формирование компетенций дисциплинам

Наименование дисциплин, формирующих компетенцию совместно	Семестры, формирования дисциплины Компетенции берутся из Учебного плана по направлению подготовки магистра			
	1	2	3	4
<i>УК-1 (Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий)</i>				
<i>Философия и методология науки в атомной энергетике [Б1.Б.4]</i>				
<i>Системная инженерия [Б1.Б.3]</i>				
<i>Подготовка к процедуре защиты и защита выпускной квалификационной работы [Б3. Д.1]</i>				
<i>Анализ больших данных [Б1.Б.6]</i>				

Таблица 3.2 – Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине			Оценочные средства	
					Текущего контроля	Промежуточной аттестации
УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	ИУК-1.2. Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению	Знать: - технологии больших данных и их применения для определения пробелов в информации, - принципы работы с большими данными.	Уметь: - разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных; - применять современные методы обработки больших данных для решения профессиональных задач	Владеть: - навыками работы с большими данными.	Выполнение и сдача 6 практических работ	Вопросы для устного собеседования – 20 вопросов

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 3 зач. ед. 108 часов, распределение часов по видам работ семестрам представлено в таблице 4.1.

Таблица 4.1 – Распределение трудоёмкости дисциплины по видам работ по семестрам для студентов очного обучения

Вид учебной работы	Трудоёмкость в час	
	Всего час.	В т.ч. по семестрам
		2 сем
Формат изучения дисциплины	с использованием элементов электронного обучения	
Общая трудоёмкость дисциплины по учебному плану	108	108
1. Контактная работа:	55	55
1.1 Аудиторная работа, в том числе:	108	108
занятия лекционного типа (Л)	34	34
занятия семинарского типа (ПЗ-семинары, практ. Занятия и др)	17	17
лабораторные работы (ЛР)	-	-
1.2 Внеаудиторная, в том числе	4	4
курсовая работа (проект) (КР/КП) (консультация, защита)	-	-
текущий контроль, консультации по дисциплине	4	4
контактная работа на промежуточном контроле (КРА)	-	-
2. Самостоятельная работа (СРС)	53	53
реферат/эссе (подготовка)	-	-
расчётно-графическая работа (РГР) (подготовка)	-	-
контрольная работа	-	-
курсовая работа/проект (КР/КП) (подготовка)	-	-
самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиум и т.д.)	53	53
Подготовка к зачету	-	-

4.2 Содержание дисциплины, структурированное по темам

Таблица 4.3 – Содержание дисциплины, структурированное по темам для студентов очного обучения

Планируемые (контролируемые) результаты освоения: код УК; ОПК; ПК и индикаторы достижения компетенций	Наименование разделов, тем	Виды учебной работы (час)					Вид СРС	Наименование используемых активных и интерактивных образовательны х технологий	Реализация в рамках Практической подготовки (трудоемкость в часах)	Наименование разработанного Электронного курса (трудоемкость в часах)	
		Контактная работа				Сам осто ятел ьна я рабо та студ енто в (час)					
		Лек ции (час)	Лаб орат ори ые рабо ты (час)	Пра кти ческ ие заня тия (час)	КСР						
2 семестр											
Раздел 1. Введение											
УК-1, ИУК-1.2	Тема 1.1. Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.	1					Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]				
	Тема 1.2 Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных	1					Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]				
	Практическая работа. Большие данные. Организация сбора и хранения больших наборов данных. Hadoop.			2			Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]		2		
	Итого по 1 разделу	2		2		-					
Раздел 2. Пакеты для работы с данными											

УК-1, ИУК-1.2	Тема 2.1. Библиотека pandas. Структуры данных в pandas, работа со структурами данных.	1				1	Подготовка к лекциям [6.1.1,6.1.2]	Разбор конкретных ситуаций		
	Тема 2.2. Библиотеки NumPy, SciPy: основные функции. Визуализация данных с matplotlib и pandas.	1,5				1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]	Разбор конкретных ситуаций		
	Тема 2.3. Облачные вычисления. Jupyter Notebook	0,5				1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]	Разбор конкретных ситуаций		
	Практическая работа. Работа с библиотеками NumPy и SciPy.			1		4	Подготовка к практической работе. [6.1.1, 6.1.2, 6.1.4]	Разбор конкретных ситуаций	1	
	Практическая работа. Работа с данными в pandas.			1			Подготовка к практической работе. [6.1.1, 6.1.2,6.1.6]	Разбор конкретных ситуаций	1	
	Итого по 2 разделу	3		2		7				
Раздел 3. Математические основы больших данных										
УК-1, ИУК-1.2	Тема 3.1. Математика в Data Science	1				1	Подготовка к лекциям [6.1.1 – 6.1.4]	Разбор конкретных ситуаций		
	Тема 3.2. Математическая статистика и элементы аналитики	2				1	Подготовка к лекциям [6.1.1 – 6.1.3]	Разбор конкретных ситуаций		
	Тема 3.3. Базовые знания по математике для работы с машинным обучением. Аппроксимация, интерполяция, функции, регрессии, матрицы и вектора.	1,5		1		1	Подготовка к лекциям [6.1.1 – 6.1.4]	Разбор конкретных ситуаций	1	

	Тема 3.4. Теория вероятностей: условная вероятность, случайные величины, теорема Байеса, распределение вероятностей.	1,5				1	Подготовка к лекциям [6.1.1 – 6.1.3]	Разбор конкретных ситуаций		
	Тема 3.5. Векторы и матрицы. Градиент.	0,5		2	1	5	Подготовка к практической работе. [6.1.1 – 6.1.4,6.1.6]		2	
	Итого по 3 разделу	6,5		3	1	9				
Раздел 4. Методы машинного обучения										
УК-1, ИУК-1.2	Тема 4.1. Примеры проектов с применением ML. Основные понятия	1				2	Подготовка к лекциям [6.1.2 – 6.1.4]	Разбор конкретных ситуаций		
	Тема 4.2. Типы, задачи машинного обучения	1				2	Подготовка к лекциям [6.1.2 – 6.1.4]	Разбор конкретных ситуаций		
	Тема 4.3. Ансамбли алгоритмов	0,5		1		5	Подготовка к практической работе. [6.1.2 – 6.1.4,6.1.6]		1	
	Тема 4.4. Оценка и улучшение качества моделей машинного обучения: перекрестная проверка, поиск по сетке, метрики качества моделей и их вычисление.	1		2		4	Подготовка к практической работе. [6.1.2 – 6.1.4,6.1.6]		2	
	Тема 4.5. Программные модули и пакеты для машинного обучения	0,5			1		Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]			
	Итого по 4 разделу	4		3	1	13				
Раздел 5. Нейронные сети										
УК-1, ИУК-1.2	Тема 5.1. Базовая теория. Отличия от классических методов машинного обучения	1				1	Подготовка к лекциям [6.1.2, 6.1.3]	Разбор конкретных ситуаций		

	Тема 5.2. Обучение с учителем. Классификация	0,5		0,5	1	Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4]	Разбор конкретных ситуаций	0,5	
	Тема 5.3. Обучение без учителя. Кластеризация	0,5		0,5	1	Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4]	Разбор конкретных ситуаций	0,5	
	Тема 5.4. Обучение с подкреплением. Нейронные сети и ансамбли	0,5			1	Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4]	Разбор конкретных ситуаций		
	Тема 5.5. Конструирование и выбор признаков. Сильно связанные признаки	1			1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]			
	Тема 5.6. Выбор типа моделей. Сравнение моделей	0,5			1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]	Разбор конкретных ситуаций		
	Тема 5.7. Работа с размерностью данных	1			1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4]	Разбор конкретных ситуаций		
	Тема 5.8. Переобучение, аугментация	1,5			1	Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4]			
	Тема 5.9. Классификация изображений	1		0,5	2	Подготовка к практической работе. [6.1.1, 6.1.2, 6.1.4,6.1.6]	Разбор конкретных ситуаций	0,5	
	Тема 5.10. Семантическая сегментация	1,5		0,5	2	Подготовка к практической работе. [6.1.1, 6.1.3, 6.1.4,6.1.6]	Разбор конкретных ситуаций	0,5	
	Тема 5.11. Задачи обработка текста	1		1	2	Подготовка к практической работе. [6.1.1, 6.1.3, 6.1.4,6.1.6]	Разбор конкретных ситуаций	1	
	Тема 5.12. Рекомендательные системы	0,5		1	1	2 Подготовка к практической работе. [6.1.1, 6.1.2, 6.1.4, 6.1.3,6.1.6]	Разбор конкретных ситуаций	1	
	Итого по 5 разделу	10,5		4	1	16			
Раздел 6. Инструменты BigData									

УК-1, ИУК-1.2	Тема 6.1. Экосистема Hadoop, Spark, Hive. Введение	1		0,5		Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4,6.1.5]	Разбор конкретных ситуаций	0,5	
	Тема 6.2. Реализации алгоритмов на MapReduce. Основы Hive.	2		0,5	4	Подготовка к лекциям [6.1.1, 6.1.3, 6.1.4,6.1.5]	Разбор конкретных ситуаций	0,5	
	Тема 6.3. Архитектура хранилищ Data Warehouse и Data Lake	1				Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4,6.1.5]			
	Тема 6.4. Airflow, Spark	1		1		Подготовка к лекциям [6.1.1, 6.1.2, 6.1.4,6.1.5,]		1	
	Тема 6.5. Промышленный Machine Learning на больших данных	3		1	1	4 Подготовка к практической работе. [6.1.1, 6.1.2, 6.1.4,6.1.5,6.1.6]	Разбор конкретных ситуаций	1	
	Итого по 6 разделу	8		3	1	8			
	Подготовка к зачету								
	Итого	34	-	17	4	53		34	

5. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.

5.1 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

1. Примерный перечень вопросов при защите практических работ:

- Что такое анализ данных? Какие технологии используются при обработке данных?
- Чем отличаются структурированные данные от неструктурированных?
- Опишите процедуру сбора и хранения больших данных?
- Какие бывают структуры данных в библиотеке pandas?
- С помощью чего визуализируются данные в python, назовите основные библиотеки?
- Что такое задача регрессии?
- Что такое градиентный спуск?
- Какие матричные преобразования используются в машинном обучении?
- Какие основные задачи решает машинное обучение?
- Что такое ансамбли алгоритмов и когда они применяются?
- Какие функции активации вам известны?
- Опишите основные способы обучения нейронных сетей? В чем их отличия от алгоритмов машинного обучения?
- Какие методы применяются для работы с данными?
- Назовите основные подходы для работы с изображениями? Какие задачи решает компьютерное зрение?
- Опишите задачу сегментации?
- Что такое обработка естественного языка, назовите основные алгоритмы?
- Назовите основной инструментарий BigData? Приведите примеры?

2. Примерный перечень вопросов для зачета:

- Основные понятия: большие данные, наука о данных, анализ данных, машинное обучение. Основные этапы извлечения знаний из данных. Примеры задач анализа данных.
- Структурированные и неструктурированные данные. Категориальные и непрерывные переменные. Методы отбора признаков (переменных). Библиотека pandas: объекты Series и DataFrame.
- Обработка данных: поиск пропущенных значений, основные методы обработки пропущенных значений, обработка пропущенных значений с помощью pandas, поиск и удаление дублирующихся значений в pandas.
- Обработка данных: описательные статистики, поиск аномалий (включая гистограммы, ящичковые диаграммы, ядерные оценки плотности), анализ выбросов и шумов. Нормализация и стандартизация данных.
- Визуализация данных: виды графиков и диаграмм. Основные инструменты визуализации данных в Python.
- Математика в Data Science: основные понятия, градиентный спуск.
- Матричные преобразования, векторы, функции регрессии.
- Кластерный анализ: иерархический кластерный анализ, построение дендрограмм, методы k-средних. Кластерный анализ в Python.
- Машинное обучение: основные понятия, задачи, которые можно решить с помощью машинного обучения. Алгоритмы машинного обучения с учителем и без учителя (перечислить). Инструменты Python, используемые в машинном обучении.

- Машинное обучение: этапы моделирования. Отбор признаков на основе модели. Оценка качества построенных моделей.
- Машинное обучение: ансамбли алгоритмов. Способы реализации ансамблей на языке python.
- Задача классификации: постановка задачи, пример моделей, понятие переобучения, оценка качества классификации, тонкая настройка модели.
- Задача регрессии: постановка задачи, пример моделей, понятие переобучения, оценка качества классификации, тонкая настройка модели.
- Нейронные сети: базовые понятия, отличия нейросетей от классических методов машинного обучения.
- Нейронные сети: обучение с учителем, задача классификации
- Нейронные сети: обучение без учителя, задача кластеризации
- Нейронные сети: обучение с подкреплением. Ансамбли нейронных сетей.
- Нейронные сети: классификация изображений, семантическая сегментация, задача обработки текста.
- Рекомендательные системы. Основные подходы к построению рекомендательных систем. Примеры рекомендательных систем.
- Инструменты BigData: основные инструменты и их применение.

Комплект оценочных средств является неотъемлемой частью ФОС и хранится на кафедре «Информатика и системы управления».

5.2 Описание показателей и критериев контроля успеваемости, описание шкал оценивания

Для оценки знаний, умений, навыков и формирования компетенции по дисциплине применяется **традиционная** система, при которой успеваемость студентов оценивается по четырехбалльной шкале «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Таблица 5.1 – Критерии оценивания результата обучения по дисциплине и шкала оценивания

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Критерии оценивания результатов обучения			
		Оценка «неудовлетворительно» / «не зачтено» 0-59% от max рейтинговой оценки контроля	Оценка «удовлетворительно» / «зачтено» 60-74% от max рейтинговой оценки контроля	Оценка «хорошо» / «зачтено» 75-89% от max рейтинговой оценки контроля	Оценка «отлично» / «зачтено» 90-100% от max рейтинговой оценки контроля
УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	ИУК-1.2. Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению	Изложение учебного материала бессистемное, неполное, отсутствует понимание особенностей и методов работы с данными, не способен применять инструменты вычислений на Python.	Имеет частичное понятие об основных особенностях и методах работы с наборами данных, частичное применение инструментов вычислений на Python при решении отдельных задач, имеет трудности в применении различных методов обучения нейронных сетей.	Знает особенности и методы работы с наборами данных, применяет на практике инструменты вычислений на Python, владеет основами работы с инструментами распределенной обработки данных	Имеет глубокие системные знания в области интеллектуального анализа данных; применяет на практике инструменты вычислений на Python и методы обучения нейронных сетей при решении отдельных задач; успешно применяет инструменты распределенной обработки данных.

Таблица 5.2 – Критерии оценивания

Оценка	Критерии оценивания
Высокий уровень «5» (отлично)	оценку « отлично » заслуживает студент, освоивший знания, умения, компетенции и теоретический материал без пробелов; выполнивший все задания, предусмотренные учебным планом на высоком качественном уровне; практические навыки профессионального применения освоенных знаний сформированы.
Средний уровень «4» (хорошо)	оценку « хорошо » заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформулировал практические навыки.
Пороговый уровень «3» (удовлетворительно)	оценку « удовлетворительно » заслуживает студент, частично с пробелами освоивший знания, умения, компетенции и теоретический материал, многие учебные задания либо не выполнил, либо они оценены числом баллов близким к минимальному, некоторые практические навыки не сформированы.
Минимальный уровень «2» (неудовлетворительно)	оценку « неудовлетворительно » заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Учебная литература

- 6.1.1. Анализ данных и процессов / А. Барсегян, М. Куприянов, И. Холод [и др.]. – 3-е изд. – Санкт-Петербург : БХВ-Петербург, 2010. – 512 с. – ISBN 978-5-9775-0368-6.
- 6.1.2. Уэс Маккинли Python и анализ данных / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 482 с.: ил. — URL: <https://e.lanbook.com/book/131721>
- 6.1.3. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : руководство / С. Рашка ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2017. — 418 с. — ISBN 978-5-97060-409-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/100905>

6.2 Справочно-библиографическая литература

- 6.1.4. Леонид Черняк. Что делать с хаосом данных? // Открытые системы.СУБД. — 2013. — № 9. — С. 16–20
- 6.1.5. Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. — Санкт-Петербург : НИУ ИТМО, 2019. — 50 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/136573>

6.3 Перечень журналов по профилю дисциплины:

- «Искусственный интеллект и принятие решений» (<http://www.aidt.ru>)
«Мир больших данных (Big data)»

6.4 Методические указания, рекомендации и другие материалы к занятиям

- 6.1.6. Методические указания для выполнения лабораторных работ по дисциплине «Анализ больших данных», для студентов направления подготовки 09.03.01 «Информатика и вычислительная техника» дневной формы обучения / НГТУ; Сост.: М.Б.Багиров, Н.Новгород, 2021

Электронные варианты всех методических указаний отправляются на электронные адреса групп.

7. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебный процесс по дисциплине обеспечен необходимым комплектом свободно распространяемого программного обеспечения (состав по дисциплине определен в настоящей РПД и подлежит обновлению при необходимости).

7.1 Перечень информационных справочных систем

Таблица 7.1 – Перечень электронных библиотечных систем

№	Наименование ЭБС	Ссылка к ЭБС
1	Лань	https://e.lanbook.com/
2	Юрайт	https://biblio-online.ru/

7.2 Перечень свободно распространяемого программного обеспечения

Таблица 7.2 – Программное обеспечение, используемое студентами очного обучения

Программное обеспечение, используемое в университете на договорной основе	Программное обеспечение свободного распространения
-	Adobe Acrobat Reader (https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader.html)
	Linux (https://www.linux.com/)
	OpenOffice (FreeWare) https://www.openoffice.org/ru/
	JDK 8 и выше (https://adoptopenjdk.net/)
	Фреймворк Java Spring 5 (https://spring.io/projects/spring-framework)
	Eclipse (https://www.eclipse.org/)
	IntelliJ Idea (https://www.jetbrains.com/ru-ru/idea/)
	git (https://git-scm.com/), github (https://github.com/)
	Maven (https://maven.apache.org/), Gradle (https://gradle.org/)
	Редактор блок-схем (https://app.diagrams.net/)

7.3 Перечень современных профессиональных баз данных и информационных справочных систем

В таблице 7.4 указан перечень профессиональных баз данных и информационных справочных систем, к которым обеспечен доступ (удаленный доступ). Данный перечень подлежит обновлению в соответствии с требованиями ФГОС ВО.

В данном разделе могут быть приведены ресурсы (ссылки на сайты), на которых можно найти полезную для курса информацию, в т.ч. статистические или справочные данные, учебные материалы, онлайн курсы и т.д.

Таблица 7.4 – Перечень современных профессиональных баз данных и информационных справочных систем

№	Наименование профессиональной базы данных, информационно-справочной системы	Доступ к ресурсу (удаленный доступ с указанием ссылки/доступ из локальной сети университета)
1	2	3
1	База данных стандартов и регламентов РОССТАНДАРТ	https://www.rst.gov.ru/portal/gost/home/standarts
2	Перечень профессиональных баз данных и информационных справочных систем	https://cyberpedia.su/21x47c0.html
3	Каталог паттернов проектирования	https://refactoring.guru/ru/design-patterns/catalog

8. ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОВЗ

В таблице 8.1 указан перечень образовательных ресурсов, имеющих формы, адаптированные к ограничениям их здоровья, а также сведения о наличии специальных технических средств обучения коллективного и индивидуального пользования. При заполнении таблицы может быть использована информация, размещенная в подразделе «Доступная среда» специализированного раздела сайта НГТУ «Сведения об образовательной организации» <https://www.nntu.ru/sveden/accenv/>

Таблица 8.1 – Образовательные ресурсы для инвалидов и лиц с ОВЗ

№	Перечень образовательных ресурсов, приспособленных для использования инвалидами и лицами с ОВЗ	Сведения о наличии специальных технических средств обучения коллективного и индивидуального пользования
1	ЭБС «Консультант студента»	озвучка книг и увеличение шрифта
2	ЭБС «Лань»	специальное мобильное приложение - синтезатор речи, который воспроизводит тексты книг и меню навигации
3	ЭБС «Юрайт»	версия для слабовидящих

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ, НЕОБХОДИМОЕ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Для контактной и самостоятельной работы обучающихся выделены помещения, оснащённые компьютерной техникой с подключением к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду организации:

- зал электронно-информационных ресурсов (ауд. 2210 – 11 компьютеров, ауд. 6119 – 9 компьютеров);
- читальный зал открытого доступа (ауд. 6162 – 2 компьютера);
- ауд. 2303, 2202, оборудованные Wi-Fi.

Перечень материально-технического обеспечения, необходимого для реализации программы бакалавриата для студентов очного обучения, включает в себя компьютерные классы.

1. Ауд. 4403 кафедры «Информатика и системы управления» - лаборатория Программирования АСО и У

Компьютеры, оснащенные необходимым оборудованием, техническими и электронными средствами обучения и контроля знаний студентов:

- 10 АРМ (терминалов);
- мультимедийный проектор Vivitek H 1180,
- экран настенный LMP 100109,
- сетевая купольная PTZ-камера AXIS M5014.

Пакеты ПО (лицензионное):

- Dr.Web (с/н H365-W77K-B5HP-N346 от 31.05.2021),
- MATLAB R2008a DVD KIT-WIN & UNIX/MAC (№ лицензии 527840, № заказа 2035235 Softline от 05.05.2008).

Пакеты ПО (распространяемое по свободной лицензии):

- Apache OpenOffice;
- Eclipse (<https://www.eclipse.org/>)
- git (<https://git-scm.com/>)

- Microsoft Visual Studio 2017 Community Edition
(<https://visualstudio.microsoft.com/ru/vs/community/>)

2. Ауд. 4408 кафедры «Информатика и системы управления» - лаборатория Информационных технологий.

Компьютеры, оснащенные необходимым оборудованием, техническими и электронными средствами обучения и контроля знаний студентов.

- 8 рабочих мест на базе тонких клиентов DellWise,
- мультимедийный проектор BenQ PB6240,
- ноутбук Lenovo V130-151KB,
- стенд для изучения автоматических систем управления на базе блока MyRio с FPGA под управлением LabView.

Пакеты ПО (лицензионное):

- Dr.Web (с/н ZNFC-CR5D-5U3U-JKGP от 20.05.2024).

Пакеты ПО (распространяемое по свободной лицензии):

- Apache OpenOffice;
- Linux Ubuntu 20.04 (<https://releases.ubuntu.com/20.04/>)
- git (<https://git-scm.com/>)
- Microsoft Visual Studio 2017 Community Edition
(<https://visualstudio.microsoft.com/ru/vs/community/>)

Также, для самостоятельной работы обучающихся выделены помещения, оснащённые компьютерной техникой с подключением к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду организации:

Таблица 9.1 – Оснащенность аудиторий и помещений для самостоятельной работы студентов по дисциплине

№	Наименование аудиторий и помещений для самостоятельной работы	Оснащенность аудиторий помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
1	1	2	3
1	6421 учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации; г. Нижний Новгород, Казанское ш., 12	Комплект демонстрационного оборудования: • ПК, с выходом на мультимедийный проектор, на базе AMD Athlon 2.8ГГц, 4 Гб ОЗУ, 250 Гб HDD, монитор 19” – 1шт. • Мультимедийный проектор Epson- 1 шт; • Экран – 1 шт.; Набор учебно-наглядных пособий	• Microsoft Windows7 (подписка DreamSpark Premium, договор №Tr113003 от 25.09.14) • Gimp 2.8 (свободное ПО, лицензия GNU GPLv3); • Microsoft Office Professional Plus 2007 (лицензия № 42470655); • OpenOffice 4.1.1 (свободное ПО, лицензия ApacheLicense 2.0) • AdobeAcrobatReader (FreeWare); • 7-zip для Windows (свободнораспространяемое ПО, лицензия GNU LGPL); - Dr.Web (с/н ZNFC-CR5D-5U3U-JKGP от 20.05.2024).

	6543 компьютерный класс - помещение для СРС, курсового проектирования (выполнения курсовых работ), г. Нижний Новгород, Казанское ш., 12)	• Проектор Accer – 1 шт; • ПК на базе IntelCoreDuo 2.93 ГГц, 2 Гб ОЗУ, 320 Гб HDD, монитор Samsung 19' – 11 шт.. ПК подключены к сети «Интернет» и обеспечивают доступ в электронную информационно-образовательную среду университета	• Microsoft Windows 7 (подписка DreamSpark Premium, договор № Tr113003 от 25.09.14); • Microsoft Office (лицензия № 43178972); • Adobe Design Premium CS 5.5.5 (лицензия № 65112135); • Adobe Acrobat Reader (FreeWare); • 7-zip для Windows (свободнораспространяемое ПО, лицензия GNU LGPL); • Dr.Web (с/н ZNFC-CR5D-5U3U-JKGP от 20.05.2024) • КонсультантПлюс (ГПД № 0332100025418000079 от 21.12.2018); Gimp 2.8 (свободное ПО, лицензия GNU GPLv3)
--	---	--	---

10. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

10.1 Общие методические рекомендации для обучающихся по освоению дисциплины, образовательные технологии

Дисциплина реализуется посредством проведения контактной работы с обучающимися (включая проведение текущего контроля успеваемости), самостоятельной работы обучающихся и промежуточной аттестации.

При преподавании дисциплины «Анализ больших данных», используются современные образовательные технологии, позволяющие повысить активность студентов при освоении материала курса и предоставить им возможность эффективно реализовать часы самостоятельной работы.

Весь лекционный материал курса сопровождается компьютерными презентациями, в которых наглядно преподносятся материал различных разделов курса и что дает возможность обсудить материал со студентами во время чтения лекций, активировать их деятельность при освоении материала. Электронные материалы лекций в период дистанционного обучения отправляются по электронной почте на адреса групп и могут быть получены до чтения лекций и проработаны студентами в ходе самостоятельной работы.

На лекциях, практических занятиях реализуются интерактивные технологии, приветствуются вопросы и обсуждения, используется личностно-ориентированный подход, технология работы в малых группах, что позволяет студентам проявить себя, получить навыки самостоятельного изучения материала, выровнять уровень знаний в группе.

Все вопросы, возникшие при самостоятельной работе над домашним заданием, подробно разбираются на практических занятиях и лекциях. Проводятся индивидуальные и групповые консультации с использованием современных информационных технологий: электронная почта, мессенджеры, Zoom.

Иницируется активность студентов, поощряется задание любых вопросов по материалу, практикуется индивидуальный ответ на вопросы студента, рекомендуются методы успешного самостоятельного усвоения материала в зависимости от уровня его базовой подготовки.

Для оценки знаний, умений, навыков и уровня сформированности компетенции применяется традиционная система контроля и оценки успеваемости студентов в процессе текущего контроля.

Промежуточная аттестация проводится в форме экзамена с учетом текущей успеваемости.

Результат обучения считается сформированным на повышенном уровне, если теоретическое содержание курса освоено полностью. При устных собеседованиях студент исчерпывающе, последовательно, четко и логически излагает учебный материал; свободно справляется с заданиями, вопросами, использует в ответе дополнительный материал. Все предусмотренные рабочей учебной программой задания выполнены в соответствии с установленными требованиями, студент способен анализировать полученные результаты, проявляет самостоятельность при выполнении заданий.

Результат обучения считается сформированным на пороговом уровне, если теоретическое содержание курса освоено полностью. При устных собеседованиях студент последовательно, четко и логически излагает учебный материал; справляется с заданиями, вопросами, требующих применения знаний; все предусмотренные рабочей учебной программой задания выполнены в соответствии с установленными требованиями, студент способен анализировать полученные результаты; проявляет самостоятельность при выполнении заданий.

Результат обучения считается несформированным, если студент при выполнении заданий не демонстрирует знаний учебного материала, допускает ошибки, неуверенно, с большими затруднениями выполняет задания, не демонстрирует необходимых умений, качество выполненных заданий не соответствует установленным требованиям, качество их выполнения оценено числом баллов ниже трех по оценочной системе, что соответствует пороговому уровню.

10.2 Методические указания для занятий лекционного типа

Лекционный курс предполагает систематизированное изложение основных вопросов тематического плана. В ходе лекционных занятий раскрываются базовые вопросы в рамках каждой темы дисциплины (Таблицы 4.4, 4.5, 4.6). Обозначаются ключевые аспекты тем, а также делаются акценты на наиболее сложные и важные положения изучаемого материала. Материалы лекций являются опорной основой для подготовки обучающихся к практическим занятиям и выполнения заданий самостоятельной работы, а также к мероприятиям текущего контроля успеваемости и промежуточной аттестации по дисциплине.

10.3 Методические указания по освоению дисциплины на лабораторных работах

Лабораторные занятия по дисциплине не предусмотрены.

10.4 Методические указания по освоению дисциплины на практических занятиях

Подготовку к каждой практической работе студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Каждая выполненная работа с оформленным отчетом подлежит защите у преподавателя.

При оценивании практических работ учитывается следующее:

- качество выполнения практической части работы и степень соответствия результатов работы заданным требованиям;
- качество оформления отчета по работе;
- качество устных ответов на контрольные вопросы при защите работы.

10.5 Методические указания по освоению дисциплины на курсовой работе

Курсовая работа не предусмотрена учебным планом.

10.6 Методические указания по самостоятельной работе обучающихся

Самостоятельная работа обеспечивает подготовку обучающегося к аудиторным

занятиям и мероприятиям текущего контроля и промежуточной аттестации по изучаемой дисциплине. Результаты этой подготовки проявляются в активности обучающегося на занятиях и в качестве выполненных практических заданий и других форм текущего контроля.

При выполнении заданий для самостоятельной работы рекомендуется проработка материалов лекций по каждой пройденной теме, а также изучение рекомендуемой литературы, представленной в Разделе 6.

В процессе самостоятельной работы при изучении дисциплины студенты могут работать на компьютере в специализированных аудиториях для самостоятельной работы, указанных в Разделе 9. В аудиториях имеется доступ через информационно-телекоммуникационную сеть «Интернет» к электронной информационно-образовательной среде университета (ЭИОС) и электронной библиотечной системе (ЭБС), где в электронном виде располагаются учебные и учебно-методические материалы, которые могут быть использованы для самостоятельной работы при изучении дисциплины.

11. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

11.1 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта в ходе контроля текущей успеваемости

Для текущего контроля знаний студентов по дисциплине проводится комплексная оценка знаний, включающая

- защиту практических работ.

11.1.1. Типовые задания для практических работ

Типовые задания для практических работ приведены в учебно-методических указаниях по проведению практических работ.

11.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта в ходе промежуточной аттестации по дисциплине

11.2.1. Защита курсового проекта/ работы

Курсовая работа не предусмотрена учебным планом.

11.2.2. Зачет для студентов очной формы обучения во 2 семестре. Проводится в виде устного собеседования по типовым вопросам.

Типовые вопросы для промежуточной аттестации в форме зачета для студентов очной формы обучения:

1. Основные понятия: большие данные, наука о данных, анализ данных, машинное обучение. Основные этапы извлечения знаний из данных. Примеры задач анализа данных.
2. Структурированные и неструктурированные данные. Категориальные и непрерывные переменные. Методы отбора признаков (переменных). Библиотека pandas: объекты Series и DataFrame.
3. Обработка данных: поиск пропущенных значений, основные методы обработки пропущенных значений, обработка пропущенных значений с помощью pandas, поиск и удаление дублирующихся значений в pandas.

4. Обработка данных: описательные статистики, поиск аномалий (включая гистограммы, ящиковые диаграммы, ядерные оценки плотности), анализ выбросов и шумов. Нормализация и стандартизация данных.
5. Визуализация данных: виды графиков и диаграмм. Основные инструменты визуализации данных в Python.
6. Математика в Data Science: основные понятия, градиентный спуск.
7. Матричные преобразования, векторы, функции регрессии.
8. Кластерный анализ: иерархический кластерный анализ, построение дендрограмм, методы k-средних. Кластерный анализ в Python.
9. Машинное обучение: основные понятия, задачи, которые можно решить с помощью машинного обучения. Алгоритмы машинного обучения с учителем и без учителя (перечислить). Инструменты Python, используемые в машинном обучении.
10. Машинное обучение: этапы моделирования. Отбор признаков на основе модели. Оценка качества построенных моделей.
11. Машинное обучение: ансамбли алгоритмов. Способы реализации ансамблей на языке python.
12. Задача классификации: постановка задачи, пример моделей, понятие переобучения, оценка качества классификации, тонкая настройка модели.
13. Задача регрессии: постановка задачи, пример моделей, понятие переобучения, оценка качества классификации, тонкая настройка модели.
14. Нейронные сети: базовые понятия, отличия нейросетей от классических методов машинного обучения.
15. Нейронные сети: обучение с учителем, задача классификации.
16. Нейронные сети: обучение без учителя, задача кластеризации.
17. Нейронные сети: обучение с подкреплением. Ансамбли нейронных сетей.
18. Нейронные сети: классификация изображений, семантическая сегментация, задача обработки текста.
19. Рекомендательные системы. Основные подходы к построению рекомендательных систем. Примеры рекомендательных систем.
20. Инструменты BigData: основные инструменты и их применение.

В полном объеме оценочные средства имеются на кафедре «Информатика и системы управления». Оценочные средства могут быть получены по требованию.