

**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 (ПИШ)

УТВЕРЖДАЮ:

Директор ПИШ:

\_\_\_\_\_ А.В. Тумасов

“ 21 ” \_\_\_\_\_ июня \_\_\_\_\_ 2023 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**M1.Б.5 Анализ больших данных**

---

(индекс и наименование дисциплины по учебному плану)

для подготовки магистров

Направления подготовки:

14.04.01 Ядерная энергетика и теплофизика

Направленность: Высокотемпературные газовые ядерные реакторные установки

14.04.02 Ядерные физика и технологии

Направленность: Ядерное топливо и основное оборудование высокотемпературных газовых реакторов

Форма обучения: очная

Год начала подготовки 2023

Выпускающая кафедра АТС

Кафедра-разработчик ИСУ

Объем дисциплины 108 / 3

часов/з.е

Промежуточная аттестация зачет с оценкой

Разработчик: Санников А.Н., ассистент

Нижний Новгород, 2023 год

Рабочая программа дисциплины «*Анализ больших данных*» разработана в соответствии с Федеральными государственными образовательными стандартами высшего образования (ФГОС ВО 3++) по направлениям подготовки:

**14.04.01 Ядерная энергетика и теплофизика;**

**14.04.02 «Ядерные физика и технологии»**

утвержденными приказами МИНОБРНАУКИ РОССИИ от:

**27.03.2018 г. № 214**

**28.02.2018 г. № 152**

на основании учебных планов, принятых УМС НГТУ им. Р.Е. Алексеева, протокол от **25.05.2023 № 22.**

Рабочая программа одобрена на заседании кафедры разработчика программы протокол от **19.06.2023г. № 35**

Зав. кафедрой к.т.н, доцент, О.П. Тимофеева \_\_\_\_\_

Программа рекомендована к утверждению ученым советом ИЯЭиТФ,  
протокол от 18 мая 2023 № 4.

Председатель УМС, директор ИЯЭиТФ \_\_\_\_\_ М.В. Легчанов

Рабочая программа зарегистрирована в УМУ, регистрационный № M1.Б.5 - ввп

Начальник МО \_\_\_\_\_ Н.Р. Булгакова  
(подпись)

Заведующая отделом комплектования НТБ \_\_\_\_\_ Н.И. Кабанина  
(подпись)

## СОДЕРЖАНИЕ

<u>1. Цели и задачи освоения дисциплины (модуля)</u> .....	4
<u>2. Место дисциплины в структуре образовательной программы</u> .....	4
<u>3. Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)</u> .....	5
<u>4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения оп во</u> .....	7
<u>5. Структура и содержание дисциплины</u> ..	<b>ОШИБКА! ЗАКЛАДКА НЕ ОПРЕДЕЛЕНА.</b>
<u>6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины</u> .....	12
<u>7. Учебно-методическое обеспечение дисциплины</u> .....	15
<u>8. Информационное обеспечение дисциплины</u> .....	16.....17
<u>9. Образовательные ресурсы для инвалидов и лиц с овз</u> .....	19
<u>10. Материально-техническое обеспечение, необходимое для осуществления образовательного процесса по дисциплине</u> .....	19
<u>11. Методические рекомендации обучающимся по освоению дисциплины</u> .....	18
<u>12. Оценочные средства для контроля освоения дисциплины</u> .....	20

## **ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)**

### **1.1 Цель освоения дисциплины**

Целью освоения дисциплины «анализ больших данных» является развитие компетенций в области разработки программных средств с использованием современных интеллектуальных технологий и математических методов обработки информации.

### **1.2 Задачи освоения дисциплины (модуля)**

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

1. Сбор, обработка, анализ и систематизация научно-технической информации по теме исследования, выбор методик и средств решения задачи.
2. Изучение автоматизированных методов сбора информации, использование методов интеллектуального анализа больших данных при помощи средств языка программирования Python.
3. Использование обработанной информации в рассматриваемой предметной области.

## **ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Учебная дисциплина «Анализ больших данных» включена в обязательный перечень дисциплин базовой части Блока1, определяющей направленность ОП. Дисциплина реализуется в соответствии с требованиями (ФГОС ВО 3++), ОП ВО и УП.

Дисциплина базируется на следующих дисциплинах в объёме 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 изучения, прежде всего, курса «Системная инженерия» [М.1.Б7], а также она имеет большое значение при подготовке к процедуре защиты и защите выпускной квалификационной работе [М.3.Д1], репрезентирующей систему компетенций магистранта.

Рабочая программа дисциплины «Анализ больших данных» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

## КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Таблица 1.- Формирование компетенций дисциплинами по направлениям подготовки:

- 14.04.01 Ядерная энергетика и теплофизика
- 14.04.02 Ядерные физика и технологии

Наименование дисциплин, формирующих компетенцию совместно	Семестры, формирования дисциплины Компетенции берутся из Учебного плана по направлению подготовки магистра»			
	1	2	3	4
<b>УК-1</b> <i>Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий</i>				
Философия и методология науки в атомной энергетике [М.1.Б4]				
Анализ больших данных [М.1.Б5]				
Системная инженерия [М.1.Б7]				
Подготовка к процедуре защиты и защита выпускной квалификационной работы [М.3. Д1]				

## **ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОП ВО**

Таблица 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	Промежуточного контроля			
<b>УК – 1</b> Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<b>ИУК-1.2.</b> Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устраниению	<b>Знать:</b> Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устраниению	<b>Уметь:</b> <ul style="list-style-type: none"> <li>- технологии больших данных и их применения для определения пробелов в информации,</li> <li>- принципы работы с большими данными.</li> </ul>	<b>Владеть:</b> <ul style="list-style-type: none"> <li>- разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных;</li> <li>- применять современные методы обработки больших данных для решения профессиональных задач</li> </ul>	Выполнение двух коллоквиумов, итогового тестирования и сдача практических работ	Зачет - 20 вопросов для устного собеседования.

## **5.1 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ**

Общая трудоёмкость дисциплины составляет 3 зач. ед. 108 часов, распределение часов по видам работ семестрам представлено в таблице 5.1.

Таблица 5.1 - 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
<b>Формат изучения дисциплины</b>	с использованием элементов электронного обучения	
<b>Общая трудоёмкость</b> дисциплины по учебному плану	<b>108</b>	<b>108</b>
<b>1. Контактная работа:</b>	<b>55</b>	<b>55</b>
<b>1.1 Аудиторная работа, в том числе:</b>	<b>108</b>	<b>108</b>
занятия лекционного типа (Л)	34	34
занятия семинарского типа (ПЗ-семинары, практ. Занятия и др)		
лабораторные работы (ЛР)	17	17
<b>1.2 Внеаудиторная, в том числе</b>	<b>4</b>	<b>4</b>
курсовая работа (проект) (КР/КП) (консультация, защита)		
текущий контроль, консультации по дисциплине	4	4
контактная работа на промежуточном контроле (КРА)		
<b>2. Самостоятельная работа (СРС)</b>	<b>53</b>	<b>53</b>
реферат/эссе (подготовка)		
расчёто-графическая работа (РГР) (подготовка)		
контрольная работа		
курсовая работа/проект (КР/КП) (подготовка)		
самостоятельное изучение разделов, самоподготовка (проработка повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиум и т.д.)	36	36
Подготовка к зачёту	17	17

## 5.2. Содержание дисциплины, структурированное по темам

### Содержание дисциплины

Таблица 5.2 - 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 образовательных технологий	Реализация в рамках Практической подготовки (трудоемкость в часах)	Наименование разработанного Электронного курса (трудоемкость в часах)				
		Контактная работа				Самостоятельная работа студентов (час)								
		Лекции (час)	Лабораторные работы (час)	Практические занятия (час)	KCP									
<b>Раздел 1. Введение в большие данные</b>														
ИУК-1.2	Тема 1.1 Введение в большие данные: основные понятия, источники Big Data.	1				1	Подготовка к лекциям [6.1.1, 6.1.2].							
	Тема 1.2 Аналитика данных. Задачи, решаемые Big Data.	1				1	Подготовка к лекциям [6.1.1, 6.1.2].	Мозговой штурм						
	<b>Итого по 1 разделу</b>	<b>2</b>				<b>2</b>								
<b>Раздел 2. Жизненный цикл аналитики данных</b>														
ИУК-1.2	Тема 2.1 Жизненный цикл аналитики данных. Понятие жизненного цикла аналитики данных. ETL(Extract, Transform, Load)-процесс.	1,5				1	Подготовка к лекциям [6.1.1, 6.1.2].							
	Тема 2.2 BusinessIntelligence (BI). Средства BI. Инструменты анализа BI.	1,5				1	Подготовка к лекциям [6.1.1, 6.1.2].	Разбор конкретных ситуаций						
	Тема 2.3 Интерактивный блокнот Jupyter на основе языка программирования	3			1	5	Подготовка к лекциям [6.1.1, 6.1.2].	Разбор конкретных ситуаций						

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 образовательных технологий	Реализация в рамках Практической подготовки (трудоемкость в часах)	Наименование разработанного Электронного курса (трудоемкость в часах)				
		Контактная работа				КСР	Самостоятельная работа студентов (час)								
		Лекции (час)	Лабораторные работы (час)	Практические занятия (час)											
	Python.														
	<b>Тема лабораторной работы:</b> “Основы синтаксиса. Поисковые алгоритмы для решения головоломок».		5					Подготовка к лабораторной работе [6.1.9]	Разбор конкретных ситуаций						
	<b>Итого по 2 разделу</b>	<b>6</b>	<b>5</b>		<b>1</b>	<b>7</b>									

### Раздел 3. Алгоритмы анализа больших данных

ИУК-1.2	<b>Тема 3.1</b> Понятие знаний в системе искусственного интеллекта. Свойства знаний. Data mining. Data analytics.	1,5			1	Подготовка к лекциям [6.1.1, 6.1.2].				
	<b>Тема 3.2</b> Интеллектуальный анализ данных, его отличия и задачи. Text Mining. Web Mining. Web Content Mining. Web Usage. Mining Social media mining.	3			2	Подготовка к лекциям [6.1.1, 6.1.2].				
	<b>Тема 3.3</b> Этапы предварительного анализа данных. Работа с DataFrame. Выбросы данных.	1,5			2	Подготовка к лекциям [6.1.1, 6.1.2].				

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 образовательных технологий	Реализация в рамках Практической подготовки (трудоемкость в часах)	Наименование разработанного Электронного курса (трудоемкость в часах)				
		Контактная работа				КСР								
		Лекции (час)	Лабораторные работы (час)	Практические занятия (час)	Самостоятельная работа студентов (час)									
	<b>Тема 3.4</b> Регрессионный анализ данных.	3			1	4	Подготовка к лекциям [6.1.3].	Разбор конкретных ситуаций						
	<b>Тема 3.5</b> Статистические методы анализа данных. Статистические гипотезы и критерии. Машинное обучение.	3				4	Подготовка к лекциям [6.1.3].							
	<b>Тема 3.6</b> Метрический и линейных классификаторы ROC-кривая. AUC-кривая.	3			1	4	Подготовка к лекциям [6.1.1, 6.1.3].							
	<b>Тема лабораторной работы:</b> “Решение задачи регрессии”		6				Подготовка к лабораторной работе [6.1.9]	Мозговой штурм						
	<b>Итого по 3 разделу</b>	<b>15</b>	<b>6</b>		<b>2</b>	<b>17</b>								
<b>Раздел 4. Классификация и кластеризация данных</b>														
ИУК-1.2	<b>Тема 4.1</b> Классификация данных. Алгоритм k-средних.	3				2	Подготовка к лекциям [6.1.3].							
	<b>Тема 4.2.</b> Кластеризация данных. Кластерный анализ.	3			1	2	Подготовка к лекциям [6.1.3].							
	<b>Тема 4.3.</b> Визуализация больших данных. Изучение библиотеки Pandas.	3,5				2		Разбор конкретных ситуаций.						
	<b>Тема лабораторной работы:</b> “Решение задачи		6				Подготовка к лабораторной работе [6.1.9]	Мозговой штурм						

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 образовательных технологий	Реализация в рамках Практической подготовки (трудоемкость в часах)	Наименование разработанного Электронного курса (трудоемкость в часах)				
		Контактная работа					Самостоятельная работа студентов (час)								
		Лекции (час)	Лабораторные работы (час)	Практические занятия (час)	КСР										
		классификации данных”													
	<b>Итого по 4 разделу</b>	<b>9,5</b>	<b>6</b>		<b>1</b>	<b>6</b>									
<b>Раздел 5. Тенденции развития больших данных</b>															
ИУК-1.2	<b>Тема 5.1</b> Основные тенденции развития методов анализа больших данных.	1,5				1	Подготовка к лекциям [6.1.1, 6.1.2, 6.1.3]	Разбор конкретных ситуаций.							
	<b>Итого по 5 разделу</b>	<b>1,5</b>				<b>1</b>									
	Подготовка к зачёту с оценкой					<b>20</b>									
	<b>Итого за семestr</b>	<b>34</b>	<b>17</b>		<b>4</b>	<b>53</b>									

## **6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

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

Для выполнения процедур оценивания составлен фонд оценочных средств, содержащий материалы для оценивания знаний, умений и навыков студентов для текущей и промежуточной аттестации.

– **Задания коллоквиума №1:**

- Понятие «Big Data» и аналитики данных;
- Цель обработки Big Data;
- Основные характеристики больших данных;
- Принцип «Трех V»;
- Задачи, решаемые Big Data;
- BI-платформа, средства BI;
- ETL-процесс;
- Оперативная аналитическая обработка данных (OLAP);
- Инструменты анализа: Power Query, MS Power BI, Pyramid Analytics;
- Понятие жизненного цикла аналитики данных;
- Синтаксис интерактивного блокнота Jupyter Notebook.

1. **Задания коллоквиума №2:**

- Регрессионный анализ данных;
- Линейная регрессия;
- Нелинейная регрессия;
- Понятие «Знание», свойства знаний;
- Основные понятия Data mining;
- Основные понятия Data analytics;
- Понятие ROC-кривая, ее свойства;
- Понятие AUC-кривая, ее свойства;
- Классификация данных, основные алгоритмы классификации.
- Кластеризация данных, основные алгоритмы кластерного анализа.

2. **Типовые вопросы текущего контроля успеваемости по дисциплине:**

1. К технологиям, используемым для сбора и обработки Big Data, можно отнести:
  - a) программное обеспечение;
  - b) оборудование;
  - c) социальные сети;
  - d) сервисные услуги;
  - e) СУБДД.
  - f) Знание
2. Data Mining — это процесс обнаружения в сырых данных...
  - a) ранее сформулированных гипотез
  - b) неочевидных закономерностей
  - c) практических закономерностей
  - d) объективных закономерностей
  - e) большого количества закономерностей

3. Большинство методов Data mining были разработаны в рамках ...
- a) классического анализа данных
  - b) теории баз данных
  - c) теории искусственного интеллекта
4. Классификация относится к стратегии:
- a) обучения с учителем
  - b) обучения без учителя
  - c) обучения с подкреплением
5. Алгоритм k-средних предназначен для решения задачи:
- a) Классификации
  - b) Кластеризации
  - c) Прогнозирования
  - d) Снижения размерности

### **3. Примерный перечень вопросов для зачета с оценкой:**

1. Понятие «Big Data» и аналитики данных. Основные характеристики больших данных.
2. Задачи, решаемые Big Data. Принцип «Трех V».
3. BI-платформа, средства BI.
4. ETL-процесс.
5. Оперативная аналитическая обработка данных.
6. Жизненный цикл аналитики данных.
7. Знание как основной инструмент искусственного интеллекта. Свойства знаний.
8. Понятия «Data mining».
9. Понятия «Data analytics».
10. Этапы предварительного анализа данных.
11. Регрессионный анализ данных. Линейная регрессия.
12. Регрессионный анализ данных. Нелинейная регрессия.
13. Статистические методы анализа данных и машинного обучения.
14. Статистические методы анализа данных. Метрический классификатор.
15. Статистические методы анализа данных. Линейный классификатор.
16. Понятия ROC-кривая и AUC-кривая. Свойства кривых.
17. Классификация данных. Алгоритм K-means.
18. Классификация данных. Алгоритм C-means.
19. Классификация данных. Алгоритм k++.
20. Кластеризация данных. Кластерный анализ.

## **6.2 Описание показателей и критериев контроля успеваемости, описание шкал оценивания**

Для оценки знаний, умений, навыков и формирования компетенции по дисциплине применяется **традиционная** система контроля и оценки успеваемости студентов.

При промежуточном контроле успеваемость студентов оценивается по пятибалльной системе «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

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

**Таблица 6 - Критерии оценивания результата обучения по дисциплине и шкала оценивания**

Код и наименование компетенции	Критерии оценивания результатов обучения			
	Оценка «неудовлетворительно» / «не засчитено» 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, вырабатывать стратегию действий	Не владеет технологиями анализа больших данных, не может сформулировать основные свойства и характеристики big data.	Владеет фрагментарными знаниями в анализе больших данных, допускает существенные ошибки в выполнении лабораторных работ, которые исправляет при помощи преподавателя, затрудняется формулировать практические результаты.	Знает материал на достаточно хорошем уровне, способен системно излагать методы анализа больших данных, при этом допускает единичные ошибки в адаптации алгоритмов к новым областям знаний.	Имеет глубокие знания всего материала структуры дисциплины, знает и применяет методы анализа больших данных, способен легко видоизменять решение задачи и аргументировать свои принимаемые решения.

<b>Оценка</b>	<b>Критерии оценивания</b>
Высокий уровень «5» (отлично)	оценку « <b>отлично</b> » 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 « <b>хорошо</b> » заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформировал практические навыки.
Пороговый уровень «3» (удовлетворительно)	оценку « <b>удовлетворительно</b> » 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 « <b>неудовлетворительно</b> » заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы.

## **7. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **7.1 Учебная литература:**

7.1.1. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джейфри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/93571>

7.1.2. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 188 с. — ISBN 978-5-507-46866-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/322664>

7.1.3. Майер-Шенбергер В., Куффер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. Language Arts & Disciplines – 2013. – 599с.

### **7.2 Справочно-библиографическая литература:**

7.2.1. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/182452>

7.2.2. Радченко, И. А. Технологии и инфраструктура Big Data : учебное пособие / И. А. Радченко, И. Н. Николаев. — Санкт-Петербург : НИУ ИТМО, 2018. — 52 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/136430>

### **7.3 Перечень журналов по профилю дисциплины:**

7.3.1. Журнал "Искусственный интеллект и принятие решений"  
(<http://aidt.ru/index.php?lang=ru>)

7.3.2. Научно-технический и научно-производственный журнал Информационные технологии [Журнал "Информационные технологии" \(novtex.ru\)](#).

7.3.3. Журнал «Информационные технологии и вычислительные системы». [Журнал «Информационные технологии и вычислительные системы» - About journal \(jitcs.ru\)](#)

## **7.4 Методические указания, рекомендации и другие материалы к занятиям**

Методические указания по выполнению лабораторных работ по дисциплине «Анализ больших данных» в бумажном варианте находятся на кафедре «Информатика и системы управления», в библиотеке НГТУ им. Р.Е. Алексеева. Электронные варианты методических указаний по выполнению лабораторных работ отправляются на электронные адреса групп.

## **8. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

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

### **8.1. Перечень информационных справочных систем**

Таблица 8. Перечень электронных библиотечных систем

<b>№</b>	<b>Наименование ЭБС</b>	<b>Ссылка к ЭБС</b>
1	Консультант студента	<a href="http://www.studentlibrary.ru/">http://www.studentlibrary.ru/</a>
2	Лань	<a href="https://e.lanbook.com/">https://e.lanbook.com/</a>
3	Юрайт	<a href="https://biblio-online.ru/">https://biblio-online.ru/</a>

### **7.2. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства необходимого для освоения дисциплины**

Таблица 9. Программное обеспечение

<b>Программное обеспечение, используемое в университете на договорной основе</b>	<b>Программное обеспечение свободного распространения</b>
-	Adobe Acrobat Reader ( <a href="https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader.html">https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader.html</a> ) Linux ( <a href="https://www.linux.com/">https://www.linux.com/</a> ) OpenOffice (FreeWare) <a href="https://www.openoffice.org/ru/">https://www.openoffice.org/ru/</a> JDK 8 и выше ( <a href="https://adoptopenjdk.net/">https://adoptopenjdk.net/</a> ) Фреймворк Java Spring 5 ( <a href="https://spring.io/projects/spring-framework">https://spring.io/projects/spring-framework</a> ) Eclipse ( <a href="https://www.eclipse.org/">https://www.eclipse.org/</a> ) IntelliJ Idea ( <a href="https://www.jetbrains.com/ru-ru/idea/">https://www.jetbrains.com/ru-ru/idea/</a> ) git ( <a href="https://git-scm.com/">https://git-scm.com/</a> ), github ( <a href="https://github.com/">https://github.com/</a> ) Maven ( <a href="https://maven.apache.org/">https://maven.apache.org/</a> ), Gradle ( <a href="https://gradle.org/">https://gradle.org/</a> ) Редактор блок-схем ( <a href="https://app.diagrams.net/">https://app.diagrams.net/</a> ) Python IDE версия 3.7 и выше

## **Перечень современных профессиональных баз данных и информационных справочных систем**

В таблице 10 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 10 - Перечень современных профессиональных баз данных и информационных справочных систем

<b>№</b>	<b>Наименование профессиональной базы данных, информационно-справочной системы</b>	<b>Доступ к ресурсу (удаленный доступ с указанием ссылки/доступ из локальной сети университета)</b>
<b>1</b>	<b>2</b>	<b>3</b>
<b>1</b>	База данных стандартов и регламентов РОССТАНДАРТ	<a href="https://www.gost.ru/portal/gost//home/standarts">https://www.gost.ru/portal/gost//home/standarts</a>
<b>2</b>	Электронная база избранных статей по философии	<a href="http://www.philosophy.ru/">http://www.philosophy.ru/</a>
<b>3</b>	Единый архив экономических и социологических данных	<a href="http://sophist.hse.ru/data_access.shtml">http://sophist.hse.ru/data_access.shtml</a>
<b>4</b>	Базы данных Национального совета по оценочной деятельности	<a href="http://www.ncva.ru">http://www.ncva.ru</a>
<b>5</b>	Справочная правовая система «КонсультантПлюс»	доступ из локальной сети
<b>6</b>	Информационно-справочная система «Техспектр»	доступ из локальной сети

## **9. ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОВЗ**

В таблице 11 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 11 - Образовательные ресурсы для инвалидов и лиц с ОВЗ

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

## **10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ, НЕОБХОДИМОЕ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ**

Учебные аудитории для проведения занятий по дисциплине, оснащены оборудованием и техническими средствами обучения, состав которых определен в данном разделе.

В таблице 11 перечислены:

- учебные аудитории для проведения учебных занятий, оснащенные оборудованием и техническими средствами обучения;

-помещения для самостоятельной работы обучающихся, которые должны оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду НГТУ.

**Таблица 11 - Оснащенность аудиторий и помещений для проведения учебных занятий и самостоятельной работы студентов по дисциплине**

<b>№</b>	<b>Наименование аудиторий и помещений для проведения учебных занятий и самостоятельной работы</b>	<b>Оснащенность аудиторий помещений и помещений</b>	<b>Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа</b>
<b>1</b>	<b>1</b>	<b>2</b>	<b>3</b>
1	<b>4409</b> учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, проведения лабораторных работ по дисциплине; г. Нижний Новгород, ул. Минина, дом 24В, корп. 4	<ul style="list-style-type: none"><li>• Мультимедийный проектор INFOCUS IN1036</li><li>• Моноблок Lenovo ThinkCentre M 72z</li><li>• Доска магнитно-маркерная Rocada 6508 эмаль 120x180см</li><li>• Экран Lumien Eco Control LEC-100115</li><li>Ноутбуки HP EliteBook 20 шт.</li></ul>	<ul style="list-style-type: none"><li>• Dr.Web (с/н S684-LRQ5-U7NH-BE97 от 11.05.22).</li><li>• Распространяемое по свободной лицензии: Apache OpenOffice</li><li>• Передаваемое ОУ на бесплатной основе в учебных целях: ОС Windows 10 (лицензия Spark для ВУЗов)</li><li>• Среда разработки: Visual Studio Community 2022</li><li>• Архиватор: 7Zip</li></ul> <p>Средство просмотра PDF: Acrobat Reader</p>

## **11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

### **11.1. 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, лабораторных 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 лабораторных занятиях и лекциях. Проводятся индивидуальные и групповые консультации с использованием современных информационных технологий: электронная почта, мессенджеры, Discord.

Инициируется 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 зачета с оценкой 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 задачами, вопросами и другими видами заданий, 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 стройно излагает учебный материал; справляется с задачами, вопросами и другими видами заданий, 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 допороговому уровню.

## 11.2 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 (Таблица 4). 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 / лабораторным работам и выполнения заданий самостоятельной работы, а также к мероприятиям текущего контроля успеваемости и промежуточной аттестации по дисциплине.

В ходе лекционных занятий рекомендуется вести конспектирование учебного материала.

### **11.3 Методические указания по освоению дисциплины на лабораторных работах**

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

При оценивании лабораторных работ учитывается следующее:

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

### **11.4. 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 (указано в таблице 11). 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.

## **12. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

### **12.1.1. Типовые задания для практических работ**

- исследовать набор данных исследуемой предметной области путем проведения корреляционного анализа;
- исследовать набор данных исследуемой предметной области путем построения ROC и AUC- кривых;
- произвести классификацию исходных данных при помощи алгоритма K-means. Метрика расстояния между объектами – расстояние Чебышева.
- произвести классификацию исходных данных при помощи модифицированного алгоритма K-means. Метрика расстояния между объектами – расстояние Чебышева.
- произвести классификацию исходных данных при помощи алгоритма K-means. Метрика расстояния между объектами – евклидово расстояние.
- произвести классификацию исходных данных при помощи алгоритма K-means. Метрика расстояния между объектами – Манхэттенское расстояние.
- произвести классификацию исходных данных алгомеративным методом. Вычисление расстояния между классами – метод Варда.

- произвести классификацию исходных данных алгомеративным методом. Вычисление расстояния между классами – метод ближайшего соседа.
- произвести классификацию исходных данных алгомеративным методом. Вычисление расстояния между классами – метод наиболее удаленного соседа.
- произвести классификацию исходных данных алгомеративным методом. Вычисление расстояния между классами – попарное среднее.
- произвести классификацию исходных данных дивизионным методом;
- произвести регрессионный анализ данных;

### **12.1.2. Типовые вопросы (задания) для устного (письменного) опроса**

- Цель обработки Big Data;
- Основные характеристики больших данных;
- Принцип «Трех V»;
- Регрессионный анализ данных;
- Понятие «Знание», свойства знаний;
- Основные понятия Data mining;
- Понятие ROC-кривая, ее свойства;
- Понятие AUC-кривая, ее свойства;
- Классификация данных, основные алгоритмы классификации.
- Кластеризация данных, основные алгоритмы кластерного анализа.

#### **Примерный тест для итогового тестирования**

##### **Вопросы для оценки компетенций УК-1:**

1. Какую часть мирового рынка Data Mining занимают услуги или консультации по эффективному внедрению этой технологии для решения актуальных бизнес-задач:

- более 75%
- около половины
- менее 10% рынка

2. Готовые алгоритмы, полная конфиденциальность информации, техническая поддержка производителя, общение с другими пользователями пакета — это преимущества использования:

- адаптация программного обеспечения под конкретную задачу
- заказ готового решения у фирмы-разработчика
- готового программного обеспечения

3. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:

- адаптированность
- не требуется дописывать программный код
- сложность подготовки данных

4. На этапе первичного исследования данных:

- всю работу осуществляет заказчик
- со стороны заказчика может потребоваться максимальное участие
- со стороны заказчика может потребоваться лишь минимальное участие

5. Постановка бизнес-задачи – это этап, который:

- формулирует конкретные бизнес-задачи, и они уже не могут быть изменены

- б) формулирует конкретные бизнес-задачи, и они могут быть изменены в ходе прохождения именно этого цикла  
в) формулирует конкретные бизнес-задачи, и они не могут быть изменены в ходе прохождения именно этого цикла

6. Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?

- a) 100Тб
- b) 1Пб
- c) 100Гб
- d) 1Тб

7. Коммерческая клиника желает установить структуру своих клиентов с точки зрения вклада в доход клиники. К какому типу относится эта задача анализа данных?

- a) прогнозирование
- b) кластеризация
- c) цензурирование
- d) классификация

8. С некоторой периодичностью персонал предприятия списывает группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Определены три категории: «ошибочные», «под сомнением», «безошибочные». К какому типу задач анализа данных относится задача о построении правила автоматического отнесения списаний к этим категориям.

- a) классификация
- b) поиск информативных признаков
- c) кластеризация
- d) цензурирование

9. К основным характеристикам Big Data относятся:

- a) Virtualization, Volume, Variability, Vehicle
- b) Variety, Velocity, Volume, Value
- c) Verification, Volume, Velocity, Visualization
- d) Video, Value, Variety, Volume

10. Медианой для выборки 1,\_3,7,10,15,16,18 является:

- a) 7,714286
- b) 7
- c) 8,5
- d) рассчитать медиану невозможно из-за пропущенных значений

11. Сколько в 1 зеттабайте тарабайт?

- a)  $1,073742 \cdot 10^9$
- b)  $2,147484 \cdot 10^9$
- c)  $1,888947 \cdot 10^7$
- d) 1024

12. Локальность данных Big Data — это:

- a) Расширение механизма обработки данных при росте объема данных
- b) Обработка и хранение происходит на одной машине
- c) Время коммуникации не может быть выше времени обработки
- d) Данные не стоит обрабатывать на сервере их хранения

13. Неверно, что к достоинствам алгоритма кластеризации k-средних относится:

- a) простота использования
- b) быстрота использования
- c) понятность и прозрачность алгоритма
- d) нечувствительность к выбросам

14. Отказоустойчивость Big Data — это, когда:

- a) Активируются до 1000 компьютеров
- b) Сбой в одном звене системы не ведет к сбоям в других звеньях
- c) Недостоверные данные удаляются из системы
- d) Данные обрабатываются на других серверах

15. Большинство методов Data mining были разработаны в рамках...

- a) теории искусственного интеллекта
- b) классического анализа данных
- c) теории баз данных

16. Дан массив >>> c = array([[1,2], [2,3], [4,5]])

Чему равен срез c[:,1]:

- a) array([1, 2, 4])
- b) array([2, 3])
- c) array([2, 3, 5])

17. Что такое регулярное выражение:

- a) шаблон, описывающий множество строк
- b) синтаксически правильное выражение на языке Python
- c) шаблон для поиска файлов в каталоге

18. Что будет получено в результате вычисления следующего выражения:

(0 < 5 <= 3) and (0 / 0):

- a) True (или 1)
- b) False (или 0)
- c) синтаксическая ошибка

19. Каким языком является Python?

- a) аспектно-ориентированный
- b) логический
- c) функциональный
- d) объектно-ориентированный

20. К какому уровню относится язык Python?

- a) к высокому
- b) к низкому
- c) к среднему
- d) нет правильного ответа

**Регламент проведения промежуточной аттестации в форме компьютерного тестирования**

Кол-во заданий в банке вопросов	Кол-во заданий, предъявляемых студенту	Время на тестирование, мин.
не менее 30 заданий	20 заданий	40 минут

Полный фонд оценочных средств для проведения промежуточной аттестации в форме компьютерного тестирования размещен в банке вопросов данного курса дисциплины в СДО eLearning Server 4G ЭИОС НГТУ.

В ходе подготовки к промежуточной аттестации обучающимся предоставляется возможность пройти тест самопроверки. Тест для самопроверки по дисциплине размещен в eLearning Server 4G ЭИОС НГТУ в свободном для студентов доступе.