



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет математики и компьютерных наук

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **Основы работы с большими данными**

Кафедра прикладной математики  
факультета математики и компьютерных наук

Образовательная программа бакалавриата

**01.03.05 - Статистика**

Направленность (профиль) программы  
*Анализ больших данных*

Форма обучения  
*Очная*

Статус дисциплины: входит в обязательную часть ОПОП

**Махачкала, 2023**

Рабочая программа дисциплины "Основы работы с большими данными" составлена в 2023 году в соответствии с требованиями ФГОС ВО по направлению подготовки - 01.03.05 Статистика (уровень бакалавриата) от 14.08.2020 г. №1032.

Разработчик: кафедра прикладной математики, Магомедова Е.С., к.ф.-м.н., доцент

Рабочая программа дисциплины одобрена:  
на заседании кафедры прикладной математики от 25.01.2023 г., протокол № 5.  
Зав. кафедрой КГ Кадиев Р.М.

на заседании Методической комиссии факультета математики и компьютерных наук от 25.01.2023 г., протокол № 4.

Председатель М.К. Ризаев Ризаев М.К.

Рабочая программа дисциплины согласована с учебно-методическим управлением  
« до » февраль 2023 г.

Начальник УМУ А.Г. Гасангаджиева Гасангаджиева А.Г.

(подпись)

## Аннотация рабочей программы дисциплины

Дисциплина "Основы работы с большими данными" входит в обязательную часть ОПОП программы *бакалавриата* по направлению подготовки 01.03.05 - Статистика.

Дисциплина реализуется на факультете *математики и компьютерных наук кафедрой прикладной математики*.

Содержание дисциплины охватывает круг вопросов, позволяющей успешно решать современные проблемы науки и техники, в обучении студентов навыкам применения фундаментальных основ современной математики и теории вероятностей для решения различных задач исследования и решения прикладных задач с использованием компьютера, в том числе для решения следующих задач: приобретение с большой степенью самостоятельности новых знаний с использованием современных образовательных и информационных технологий, основные методы, способы и средства получения, хранения, переработки информации, умение использовать для решения коммуникативных задач современные технические средства и информационные технологии с использованием традиционных носителей информации, распределенных баз знаний, а также информации в глобальных компьютерных сетях, моделирование технических объектов и технологических процессов с использованием стандартных пакетов и средств автоматизированного проектирования, готовность проводить эксперименты по заданным методикам с обработкой и анализом результатов

Дисциплина нацелена на формирование следующих компетенций выпускника: профессиональных – ОПК-1, ОПК-2, ПК-6, ПК-7.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия, самостоятельная работа.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости: контрольная работа, коллоквиум и промежуточный контроль в форме экзамена.

Объем дисциплины 3 зачетных единиц (108 часов), в том числе в академических часах по видам учебных занятий:

Семестр	Учебные занятия						Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)	
	в том числе							
	Всего	Контактная работа обучающихся с преподавателем						СРС, в том числе экзамен
		Лекции	Лабораторные занятия	Практические занятия	КСР	Консультации		
2	108	14		14			80	Экзамен
Итого	108	14		14			80	

### 1. Цели освоения дисциплины.

*Целями* освоения дисциплины "Основы работы с большими данными" является формирование и приобретение способностей у будущих специалистов :

- осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач на основе применения больших данных;
- осуществлять сбор, обработку и статистический анализ данных, необходимых для решения поставленных экономических задач с применением методов анализа и обработки больших данных;
- понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности с использованием технологий больших данных.

*Задачи курса:*

Задачи дисциплины «Основы работы с большими данными» состоят в

приобретении навыков, знаний и умений для достижения следующих индикаторов выполнения требований по освоению дисциплины:

- Осуществляет поиск необходимой информации, хранящейся в структуре больших данных, опираясь на результаты анализа поставленной задачи;
- Разрабатывать варианты решения проблемной ситуации на основе критического анализа доступных источников информации в структуре больших данных;
- Использовать основные методы, средства получения, представления, хранения и обработки статистических данных с использованием методов и технологий больших данных;
- Применять статистические методы обработки собранных данных, использует анализ данных, необходимых для решения поставленных экономических задач на основе больших данных;
- Использовать соответствующие содержанию профессиональных задач современные цифровые информационные технологии, основываясь на принципах их работы, в том числе, на принципах обработки больших данных;
- Понимать принципы работы современных цифровых информационных технологий, соответствующих содержанию профессиональных задач на основе методов и принципов хранения, выборки и обработки больших данных.

## 2. Место дисциплины в структуре ОПОП бакалавриата.

Дисциплина " Основы работы с большими данными " входит в обязательную часть ОПОП программы *бакалавриата* по направлению подготовки 01.03.05 - Статистика. Знания работы с большими данными необходимы для данной специальности, как для освоения различных смежных дисциплин, так и крайне нужны для создания математических моделей социально-экономических процессов с целью прогнозирования показателей этих процессов.

Изучение учебной дисциплины ««Основы работы с большими данными» является базовым для последующего освоения программного материала учебных дисциплин: «Методы научных исследований».

Знания, приобретенные при изучении дисциплины "Основы работы с большими данными", могут найти применение при выполнении индивидуальных заданий, курсовом и дипломном проектировании.

## 3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения).

Код и наименование компетенции ОПОП	и из	Код и наименование индикатора достижения компетенций (в соответствии с ОПОП)	Планируемые результаты обучения	Процедура освоения
<b>ОПК-1.</b> Способен осуществлять статистическое наблюдение с использованием стандартных методик и технических средств, включая формирование выборочной совокупности	и	<b>ОПК-1.1.</b> Знает источники, основные способы сбора, поиска и систематизации статистической информации.	Знает: стандартные методы и технические средства для статистических наблюдений. Умеет: применить стандартные методы и технические средства при статистических наблюдений. Владеет: методами и техническими средствами для статистических	Контрольные работы, экзамен

подготовку статистического инструментария		наблюдений.	
	<b>ОПК-1.2.</b> Умеет собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации.	Знает: собирать данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации. Умеет: собирать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации. Владеет: методами сбора данных об объекте исследования и выбора соответствующий инструментарий для обработки информации.	
	<b>ОПК-1.3.</b> Владеет статистическими методами обработки информации, в том числе с применением информационно-коммуникационных технологий.	Знает: статистические методы обработки информации, в том числе с применением информационно-коммуникационных технологий. Умеет: применять статистические методы для обработки информации, в том числе с применением информационно-коммуникационных технологий. Владеет: статистическими методами обработки информации, в том числе с применением информационно-коммуникационных технологий	
<b>ОПК-2.</b> Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том	<b>ОПК- 2.1.</b> Знает основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач.	Знает: основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач. Умеет: применять основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач.	Контрольные работы, экзамен

<p>числе применением необходимой вычислительной техники стандартных компьютерных программ</p>	с		<p>Владеет: основными моделями решения функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач.</p>
	и	<p><b>ОПК-2.2.</b> Умеет формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач.</p>	<p>Знает: формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач. Умеет: формировать массивы сводной статистической информации, применять математические и статистические методы при решении типовых профессиональных задач. Владеет: математическими и статистическими методами для формирования массивов сводной статистической информации для решения типовых профессиональных задач.</p>
		<p><b>ОПК-2.3.</b> Владеет навыками применения математического и статистического инструментария для решения прикладных задач, методами работы с современной вычислительной техникой . .</p>	<p>Знает: как применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. Умеет: применить математические и статистические инструментария и современную вычислительную технику для решения прикладных задач. Владеет: навыками применения математического и статистического инструментария для</p>

		решения прикладных задач, методами работы с современной вычислительной техникой.	
<b>ПК-6.</b> Способен формировать источники больших данных	<b>ПК-6.1.</b> Знает формировать источники больших данных.	Знает: формировать источники больших данных. Умеет: формировать источники больших данных. Владеет: навыками формирования источников больших данных.	Коллоквиум, экзамен
	<b>ПК-6.2.</b> Умеет формировать источники больших данных.	Знает: формировать источники больших данных. Умеет: формировать источники больших данных. Владеет: навыками формирования источников больших данных.	
	<b>ПК-6.3.</b> Владеет навыками формирования источников больших данных.	Знает: формировать источники больших данных. Умеет: формировать источники больших данных. Владеет: навыками формирования источников больших данных.	
<b>ПК-7.</b> Способен осуществлять выбор методов и инструментальных средств анализа	<b>ПК-7.1.</b> Знает осуществлять выбор методов и инструментальных средств анализа больших данных	Знает: осуществлять выбор методов и инструментальных средств анализа больших данных для решения	Контрольные работы, экзамен

больших данных для решения профессиональных задач.	для решения профессиональных задач.	профессиональных задач. Умеет: осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач.	
	<b>ПК-7.2.</b> Умеет осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач.	Владеет: навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.	
	<b>ПК-7.3.</b> Владеет навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.		

#### 4.Объем, структура и содержание дисциплины.

4.1 .Объем дисциплины составляет 3 зачетных единицы, 108 академических часа.

4.2 . Структура и содержание дисциплины (модули).

##### 4.2.1. Структура дисциплины.

№ п /	Название разделов и тем дисциплины	Семестр	Аудиторные занятия, в том числе				Самост. работа	Общ	Формы текущего контроля успеваемости (по неделям семестра)
			Лекции	Практ. занятия	Лаборат. работы	Контр.сам. раб.			
<b>Модуль 1.</b>		<b>Определение Больших данных. Метаданные</b>							
1.	Предмет, задачи, критерии и принципы Больших данных (БД). Создание данных (Data Generation/Data Capture)	2	2	2			4	8	Контрольная работа, коллоквиум.
2.	Обслуживание данных (Data Maintenance) . Синтез данных (Data Synthesis).Использование данных (Data Usage) . Публикация данных (Data Publication) Архивация данных . (Data Archival). Уничтожение данных (Data Purging)	2	2	2			6	10	
3.	<b>Метаданные</b> Понятие метаданных. Жизненный цикл метаданных. Оценка требований и анализ контента. Спецификация системных	2	2	2			6	10	



	требований. Система метаданных. Сервис и оценка								
4.	<b>Большие данные. Системы управления Большими данными</b> Распределенные файловые системы. Распределенные фреймворки. Бенчмаркинг. Серверное программирование. Планирование. Системы развертывания. Интеграция данных. Информационная безопасность. Машинное обучение. Базы данных NoSQL и новые SQL базы данных	2	2	2			4	8	
Итого по первому модулю.			8	8			20	36	
<b>Модуль 2.</b>		<b>Параллельные Алгоритмы. Программные платформы</b>							
5	Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer). Представление результатов (Consumption Layer)	2	2	2			8	12	
6.	Параллельные алгоритмы для работы с данными. Операторы Map и Reduce Оператор Reduce (свертка) Оператор Map Лямбда-архитектура	2	2	2			8	12	
7.	Программные платформы и системы для Больших данных. Системы управления потоками данных. Системы хранения Больших данных. Платформы Больших данных. Обработка данных в реальном времени. Системы управления Большими данными. Аналитические платформы	2	2	2			8	12	
Итого по второму модулю			6	6			24	36	
<b>Модуль 3. Подготовка к экзамену. Экзамен</b>								36	
<b>Итого за II семестр</b>			<b>12</b>	<b>16</b>	<b>16</b>		<b>44</b>	<b>108</b>	

#### 4.3. Содержание дисциплины, структурирование по темам (разделом)

### 4.3.1. Содержание лекционных занятий по дисциплинам.

#### **Модуль 1. Определение Больших данных. Метаданные.**

**Тема 1.** Предмет, задачи, критерии и принципы Больших данных (БД).  
Создание данных (Data Generation/Data Capture)

**Тема 2.** Обслуживание данных (Data Maintenance) . Синтез данных (Data Synthesis).Использование данных (Data Usage) . Публикация данных (Data Publication) Архивация данных . (Data Archival). Уничтожение данных (Data Purging)

#### **Тема 3. Метаданные**

Понятие метаданных. Жизненный цикл метаданных. Оценка требований и анализ контента. Спецификация системных требований. Система метаданных. Сервис и оценка

#### **Тема 4. Большие данные. Системы управления Большими данными**

Распределенные файловые системы. Распределенные фреймворки. Бенчмаркинг. Серверное программирование. Планирование. Системы развертывания. Интеграция данных. Информационная безопасность. Машинное обучение. Базы данных NoSQL и новые SQL базы данных.

#### **Модуль 2. Параллельные Алгоритмы. Программные платформы переменных.**

##### **Тема 5.**

Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer).  
Представление результатов (Consumption Layer)

**Тема 6.** Параллельные алгоритмы для работы с данными. Операторы Map и Reduce  
Оператор Reduce (свертка). Оператор Map. Лямбда-архитектура

##### **Тема 7. Программные платформы и системы для Больших данных.**

Системы управления потоками данных. Системы хранения Больших данных.

Платформы Больших данных. Обработка данных в реальном времени. Системы управления Большими данными. Аналитические платформы

### 4.3.2. Содержание практических занятий по дисциплине.

#### **Модуль 1. Определение Больших данных. Метаданные.**

**Тема 1.** Предмет, задачи, критерии и принципы Больших данных (БД).  
Создание данных (Data Generation/Data Capture)

**Тема 2.** Обслуживание данных (Data Maintenance) . Синтез данных (Data Synthesis).Использование данных (Data Usage) . Публикация данных (Data Publication) Архивация данных . (Data Archival). Уничтожение данных (Data Purging)

#### **Тема 3. Метаданные**

Понятие метаданных. Жизненный цикл метаданных. Оценка требований и анализ контента. Спецификация системных требований. Система метаданных. Сервис и оценка

#### **Тема 4. Большие данные. Системы управления Большими данными**

Распределенные файловые системы. Распределенные фреймворки. Бенчмаркинг. Серверное программирование. Планирование. Системы развертывания. Интеграция данных. Информационная безопасность. Машинное обучение. Базы данных NoSQL и новые SQL базы данных.

#### **Модуль 2. Параллельные Алгоритмы. Программные платформы переменных.**

##### **Тема 5.**

Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer).

Представление результатов (Consumption Layer)

#### **Тема 6.**

Параллельные алгоритмы для работы с данными. Операторы Map и Reduce  
Оператор Reduce (свертка). Оператор Map. Лямбда-архитектура

#### **Тема 7.**

Программные платформы и системы для Больших данных.  
Системы управления потоками данных. Системы хранения Больших данных.  
Платформы Больших данных. Обработка данных в реальном времени.  
Системы управления Большими данными. Аналитические платформы

#### **4.3.3. Содержание лабораторных занятий по дисциплине.**

*Не предусмотрено*

### **5. Образовательные технологии.**

В основе преподавания дисциплины «*Основы работы с большими данными*» лежит лекционно-семинарская система обучения в сочетании с самостоятельным изучением определенных разделов. Это связано с необходимостью активного продумывания теоретического материала и дальнейшего приложения его к прикладным задачам. Индивидуальные особенности обучающихся учитываются подбором заданий разного уровня сложности для самостоятельной работы студентов.

По данной дисциплине учебным планом предусмотрено также проведение занятий в интерактивных формах. Лекции проводятся в аудиториях, оснащенных видеопроекторами. В университете функционирует Центр современных образовательных технологий, в котором предусматриваются мастер-классы специалистов.

### **6. Учебно-методическое обеспечение самостоятельной работы студентов.**

#### **6.1. Виды и порядок выполнения самостоятельной работы.**

Самостоятельная работа студентов (СРС) включает контролируемую и внеаудиторную самостоятельную работу, направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки. Часть программного материала выносится для самостоятельного внеаудиторного изучения с последующим текущим или итоговым контролем знаний на занятиях или экзамене. Контроль СРС и оценка ее результатов организуется как самоконтроль (самооценка) студента, а также как контроль и оценка со стороны преподавателя, например в ходе собеседования. Баллы, полученные по СРС студентом, обязательно учитываются при итоговой аттестации по курсу.

Формы контроля СРС включают: тестирование; устную беседу по теме с преподавателем; выполнение индивидуального задания и др. Роль студента в СРС - самостоятельно организовывать свою учебную работу по предложенному преподавателем, методически обеспеченному плану. СРС по курсу учитывает индивидуальные особенности слушателей и включает не только задания, связанные с решением типовых задач, но также творческие задания, требующие самостоятельно «добывать» знания из разных областей, группировать и концентрировать их в контексте конкретной решаемой задачи. Технология обучения предусматривает выработку навыков презентации результатов выполненного индивидуального задания и создание условий для командной работы над комплексной темой с распределением функций и ответственности между членами коллектива. Оценка результатов выполнения индивидуального задания осуществляется по критериям, известным студентам, отражающим наиболее значимые аспекты контроля за выполнением этого вида работ.

#### **6.2. Темы и содержание самостоятельной работы.**

1. Понятие о больших данных
2. Программа Excel.
3. Программа Statistica.
4. SPSS.
5. Другие программы.
6. Методы оцифровки
7. Реализация статистических методов в пакетах прикладных программ
8. Формы представления данных: таблицы сопряженности разного вида, кодирование.
9. Методы обработки данных. Навыки работы со статистическими таблицами.
10. Понятие о технологии Data Mining. Реализация в пакетах прикладных программ.
11. Сетевые технологии Data Mining.
12. Примеры применения в социологии и экономике.
13. Реляционные базы данных. Параллельные базы данных.
14. Новые технологии обработки и хранения больших данных
15. Распределённые файловые системы.
16. NoSQL СУБД.
17. Технология Map-Reduce.
18. GOOGLE BIGTABLE.
19. MapReduce.
20. Технологии поиска. Обычный поиск.
21. Полнотекстовый поиск. Параллельные запросы.
22. Интеграция данных из различных источников.
23. Технология поиска и интеграции. Программные средства.
24. ETL процесс по обработке отчётов

Разделы и темы для самостоятельного изучения.	Виды и содержание самостоятельной работы.
<p>1. Реализация статистических методов в пакетах прикладных программ</p> <p>Формы представления данных: таблицы сопряженности разного вида, кодирование.</p>	<p>-конспектирование первоисточников и другой учебной литературы;</p> <p>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</p> <p>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</p> <p>-работа с тестами и вопросами для самопроверки.</p>
<p>2. Примеры применения в социологии и экономике.</p>	<p>-конспектирование первоисточников и другой учебной литературы;</p> <p>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</p> <p>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</p> <p>-работа с тестами и вопросами для самопроверки.</p>

3. Понятие о технологии Data Mining. Реализация в пакетах прикладных программ.	Изучение документации программного обеспечения.
4. GOOGLE BIGTABLE	Изучение документации программного обеспечения.
5. Интеграция данных из различных источников.	-конспектирование первоисточников и другой учебной литературы; -проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях; -поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору; -работа с тестами и вопросами для самопроверки.
6. . ETL процесс по обработке отчётов	-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору; - работа с тестами и вопросами для самопроверки.
7. Роль и сущность многомерных статистических методов в экономике, управлении, финансах, социальных науках: постановка основных задач, примеры практического использования в социально-экономических исследованиях.	-конспектирование первоисточников и другой учебной литературы; -проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях; -поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору, -работа с тестами и вопросами для самопроверки.

**7. Фонд оценочных средств, для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.**

**7.1. Типовые контрольные задания.**

**7.1.1. Перечень вопросов к экзамену:**

1. Основные методы критического анализа и основы системного подхода как общенаучного метода изучения больших данных
2. Анализ задач больших данных с использованием основ критического анализа и системного подхода
3. Поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации на основе больших данных
4. Жизненный цикл больших данных
5. Критерии сопоставления различных вариантов решения поставленной задачи создания больших данных (Data Generation/Data Capture)
6. Критический анализ собранной информации на соответствие ее условиям и

критериям решения поставленной задачи при обслуживании больших данных (Data Maintenance)

7. Отличие фактов от мнений, интерпретаций и оценок при анализе собранной информации в процессе синтеза больших данных (Data Synthesis)

8. Сопоставление и оценка различных вариантов решения поставленной задачи, определяя их достоинства и недостатки, при использовании больших данных (Data Usage)

9. Работа с национальными и международными базами данных с целью поиска информации, необходимой для решения поставленных экономических задач с использованием методов больших данных

10. Расчет экономических и социально-экономических показателей, характеризующих деятельность хозяйствующих субъектов на основе типовых методик и действующей нормативно-правовой базы с использованием методов больших данных

11. Представление наглядной визуализации больших данных (публикация больших данных) (Data Publication)

12. Архивация больших данных (Data Archival)

13. Уничтожение больших данных (Data Purging)

14. Понятие метаданных в больших данных

15. Жизненный цикл метаданных в больших данных

16. Оценка требований и анализ контента в больших данных

17. Спецификация системных требований в больших данных

18. Система метаданных в больших данных

19. Сервис и оценка в больших данных

20. Распределенные файловые системы больших данных

21. Распределенные фреймворки

22. Бенчмаркинг в больших данных

23. Серверное программирование в больших данных

24. Планирование больших данных с использованием основ теории вероятностей, математической статистики и эконометрики, в том числе, методов и форм организации статистического наблюдения, методологии первичной обработки статистической информации

25. Планирование больших данных в части экономической информации с использованием следующих методов анализа и обработки: временные ряды, перекрёстные (cross-section) данные, панельные данные; основы регрессионного анализа (линейная модель множественной регрессии); суть метода наименьших квадратов (МНК) и его применение в экономическом анализе; основные методы диагностики (проверки качества) эконометрических моделей

26. Системы развертывания больших данных

27. Интеграция больших данных

28. Информационная безопасность больших данных

29. Машинное обучение на основе больших данных

30. Базы данных NoSQL и новые SQL базы данных

31. Прием больших данных (Data Ingestion)

32. Сбор больших данных (Data Staging)

33. Анализ больших данных (Analysis Layer)

34. Представление результатов больших данных (Consumption Layer)

35. Операторы Map и Reduce в больших данных

36. Оператор Reduce (свертка) в больших данных

37. Оператор Map в больших данных

38. Лямбда-архитектура в больших данных

39. Системы управления потоками данных в системах больших данных

40. Системы хранения больших данных

41. Платформы больших данных

42. Проведение статистических тестов и построение доверительных интервалов, определение статистических свойств полученных оценок при обработке больших данных в реальном масштабе времени
43. Системы управления большими данными
44. Проведение анализа и содержательной интерпретации полученных результатов на основе больших данных с использованием инструментария аналитических платформ

### 7.1.2. Примерные контрольные вопросы к коллоквиумам.

1. Понятие и классификация больших данных. Размерность и объем данных. «Проклятие размерности».
2. Понятие и формирование озера данных Data Lake.
3. Методология исследования данных CRISP-DM. Задачи каждой фазы и возможные ошибки.
4. Методология исследования данных SEMMA. Отличия от CRISP-DM.
5. Компетенции и состав команды для работы с данными.
6. Правовые аспекты организации защиты персональных данных регламент защиты персональных данных (GDPR).
7. Задачи и методы извлечения знаний из больших данных.
8. Методы сбора данных из различных источников.
9. Методы предобработки данных.
10. Анализ социальных сетей.
11. Визуализация данных.
12. Назначение и характеристика компонент экосистемы Hadoop.
13. Парадигма MapReduce.
14. Apache Spark и его компоненты.
15. Облачные платформы для работы с большими данными.
16. Масштабирование и многоуровневое хранение данных: Парадигма NoSQL.
17. Особенности представления и обработки символьной и численной информации в нейронных сетях.
18. Методы и средства аннотирования данных для обучения нейронных сетей.
19. Современные алгоритмы и принципы обучения нейронных сетей, особенности обучения сетей различной структуры
20. Методы оценки качества нейросетевых моделей.
21. Валидация и интерпретируемость результатов работы нейросетевых моделей.
22. Системы нечеткого вывода: нечеткие высказывания, вывод в нечеткой логике, правила нечетких продукций.
23. Понятие нечеткой нейронной сети: определение и виды архитектур.
24. Обучение и алгоритмы оптимизации нечеткой нейронной сети.
25. Центры обработки данных: назначение, оборудование, уровни.

### 7.1.3. Задания для решения кейса/комплект:

#### Кейс 1. Подготовка плана проекта по анализу данных

Требуется представить план проекта по анализу данных на заданную тему (например, прогнозирование оттока клиентов фирмы).

Работа состоит из следующих этапов:

- определение целей проекта и требований со стороны заказчика;
- определение цели анализа данных;
- разработка плана проекта;
- распределение обязанностей участникам проекта в зависимости от выполняемой роли.

#### Методы и технологии анализа больших данных

#### Практические занятия

1. Методы сбора открытых данных из сети интернет
2. Разведочный анализ и визуализация данных
3. Разработка предсказательной модели

#### 4. Анализ поведения пользователей в сети интернет

##### **Кейс 2. Разработка требований заказчика к информационно-аналитической системе**

Цель: освоение базовых инструментов составления требований заказчика.

Требуется описать в 3-5 пользовательских историях и 1-2 сценариях использования любое приложение на ноутбуке или мобильном телефоне. Приветствуется добавление нового функционала.

##### **Кейс 3. Проведение переговоров с подрядчиками и поставщиками оборудования и программного обеспечения**

Требуется определить требования и составить план переговоров с подрядчиками и поставщиками оборудования и программного обеспечения

###### **Тематика докладов с презентациями:**

1. Ассоциативные правила. Поиск ассоциативных правил.
2. Кластеризация. Алгоритм кластеризации k-means.
3. Прогнозирование с помощью линейной регрессии.
4. Классификация с помощью нейросети.
5. Классификация с помощью деревьев решений.
6. Программы статистической обработки информации.
7. Прогнозирование и предвидение в социально-политических и медиа процессах

###### **Обучение на больших данных**

###### **Практические работы**

1. Проектирование и обучение глубокой нейронной сети по распознаванию образов
2. Разработка нечетко-логического регулятора многомерного процесса



#### **7.1. 4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

##### **Фонды оценочных средств по дисциплине**

Фонд оценочных средств позволяет оценить знания, умения и уровень приобретенных компетенций.

Фонд оценочных средств по дисциплине включает:

- вопросы к экзамену;
- комплект заданий к практическим работам.

#### **7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.**

##### **а) Критерии оценивания компетенций (результатов).**

- оценки "отлично" заслуживает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоивший основную литературу и знакомый с дополнительной литературой, рекомендованной программой. Как правило, оценка "отлично" выставляется студентам, усвоившим взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявившим творческие способности в понимании, изложении и использовании учебно-программного материала.

- оценки "хорошо" заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в программе задания, усвоивший основную литературу, рекомендованную в программе. Как правило, оценка "хорошо" выставляется студентам, показавшим систематический характер знаний по дисциплине и способным к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.

-оценки "удовлетворительно" заслуживает студент, обнаруживший знания основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справляющийся с выполнением заданий, предусмотренных программой, знакомый с основной литературой, рекомендованной программой. Как правило, оценка "удовлетворительно" выставляется студентам, допустившим погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.

-оценка "неудовлетворительно" выставляется студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных программой заданий. Как правило, оценка "неудовлетворительно" ставится студентам, которые не могут продолжить обучение или приступить к профессиональной деятельности по окончании вуза без дополнительных занятий по соответствующей дисциплине.

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля -50%.

Текущий контроль по дисциплине включает:

- посещение занятий – 10 баллов,
- участие на практических занятиях – 10 баллов
- выполнение лабораторных заданий – 10 баллов,
- коллоквиум – 30 баллов,
- выполнение аудиторных контрольных работ – 40 баллов.

Промежуточный контроль по дисциплине включает:

- устный опрос(экзамен)- 100 баллов.

##### **Рекомендуемые границы оценок:**

- «отлично» - не менее 86% правильных ответов,
- «хорошо» - 66-85% правильных ответов,
- «удовлетворительно» - 51-65% правильных ответов,
- «неудовлетворительно» - менее 50% правильных ответов.

## 8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

### III. УЧЕБНО-МЕТОДИЧЕСКОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

#### Основная литература:

1. Агальцов, В. П. Базы данных: в 2 кн. Книга 2. Распределенные и удаленные базы данных: учебник / В.П. Агальцов. - Москва: ИД «ФОРУМ»: ИНФРА-М, 2018. - 271 с. - (Высшее образование: Бакалавриат). - ISBN 978-5-16-105263-1. - Текст: электронный. – Режим доступа: <https://znanium.com/catalog/product/929256>
2. Гордеев, С. И. Организация баз данных в 2 ч. Часть 1 : учебник для вузов / С. И. Гордеев, В. Н. Волошина. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 310 с. — (Университеты России). — ISBN 978-5-534-04469-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/437731>

#### Дополнительная литература:

1. Крутиков, В.Н. Анализ данных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово: Кемеровский государственный университет, 2014. - 138 с.: ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>
2. Жуковский, О.И. Информационные технологии и анализ данных: учебное пособие/ О.И. Жуковский; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск: Эль Контент, 2014. - 130 с. : схем., ил. - Библиогр.: с. 126. - ISBN 978-5-4332-0158-3; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=480500>
3. Базы данных в высокопроизводительных информационных системах: учебное пособие / авт.-сост. Е.И. Николаев; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2016. - 163 с.: ил. - Библиогр.: с.161.; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=466799>.
4. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб: Университет ИТМО, 2018. – 52 с.
5. Туманов, В.Е. Проектирование хранилищ данных для систем бизнес-аналитики: учебное пособие / В.Е. Туманов. - Москва : Интернет-Университет Информационных Технологий, 2010. - 616 с.: ил., табл., схем. - (Основы информационных технологий). - ISBN 978-5-9963-0353-3; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=233492>
6. Добронез, Б.С. Численный вероятностный анализ неопределенных данных: монография/ Б.С. Добронез, О.А. Попова; Министерство образования и науки Российской Федерации, Сибирский Федеральный университет. - Красноярск: Сибирский федеральный университет, 2014. - 168 с.: граф., ил. - Библиогр. в кн. - ISBN 978-5-7638-3093-4; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&>

#### Нормативные правовые документы:

1. Федеральный закон "Об информации, информационных технологиях и о защите информации" от 27.07.2006 N 149-ФЗ.

11

2. ГОСТ 19.201-78 ЕСПД. Межгосударственный стандарт. Единая система программной документации. Техническое задание. Требования к содержанию и

оформлению.

#### **перечень информационно-справочных систем**

1. <http://www.consultant.ru> - Справочно-правовая система Консультант Плюс;
2. <http://www.garant.ru> - Справочно-правовая система Гарант.

#### **перечень электронно-образовательных ресурсов**

1. <https://eios.vfreu.ru> - ЭИОС Воронежского филиала РЭУ им. Г.В. Плеханова
2. <https://znanium.com> - ЭБС «Znanium. com»
3. <https://urait.ru> - ЭБС «Юрайт»
4. <https://biblioclub.ru> - ЭБС «Университетская библиотека online»
5. <https://www.iprbookshop.ru/> - ЭБС IPRbooks
6. <https://elibrary.ru> - НЭБ «eLIBRARY»

#### **перечень профессиональных баз данных**

1. Информационная система «Единое окно доступа к образовательным ресурсам. Раздел. Информатика и информационные технологии» - [http://window.edu.ru/catalog/?p\\_rubr=2.2.75.6](http://window.edu.ru/catalog/?p_rubr=2.2.75.6)
2. <https://habr.com/ru/all/> - система тематических коллективных блогов, связанных с информационными технологиями, бизнесом и интернетом
3. База данных Минэкономразвития РФ «Информационные системы Министерства в сети Интернет» - <http://economy.gov.ru/minec/about/systems/infosystems/>

#### **перечень ресурсов информационно-телекоммуникационной сети "интернет", необходимых для освоения дисциплины**

1. <https://intuit.ru/> - Интернет-Университет Информационных Технологий (содержит доступ к бесплатным учебным курсам по информационным технологиям и системам, учебную и методическую литературу)
2. <https://colab.research.google.com/> - веб-сервис для разработчиков на языке Python.
3. <http://amursu.ru> (amursu.ru)
4. <http://www.biblioclub.ru> (biblioclub.ru)
5. <http://www.exponenta.ru> (exponenta.ru)
6. Интернет - библиотека образовательных изданий - <http://www.iqlib.ru>

#### **перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения**

1. Операционная система Microsoft Windows 7 и старше.
2. Пакет прикладных программ Microsoft Office 2010 и старше.
3. Система управления базами данных PostgreSQL.
4. Яндекс.Браузер\_\_

### **10. Методические указания для обучающихся по освоению дисциплины.**

Для изучения теоретического курса студентам необходимо использовать лекционный материал, учебники и учебные пособия из списка основной и дополнительной литературы, интернет источники.

По дисциплине «Основы работы с большими данными» в конце каждого модуля проводится контрольная работа.

В контрольную работу включаются теоретические вопросы и задачи тех типов, которые были разобраны на предшествующих практических занятиях.

Рабочей программой дисциплины «Основы работы с большими данными» предусмотрена самостоятельная работа студентов в объеме 44 часов. Самостоятельная работа проводится с целью углубления знаний по дисциплине и предусматривает:

- чтение студентами рекомендованной литературы и усвоение теоретического материала дисциплины;
- подготовку к практическим занятиям;
- выполнение индивидуальных заданий;
- подготовку к контрольным работам и экзамену.

С самого начала изучения дисциплины студент должен четко уяснить, что без систематической самостоятельной работы успех невозможен. Эта работа должна регулярно начинаться сразу после лекционных и практических занятий, для закрепления только что пройденного материала.

После усвоение теоретического материала можно приступить к самостоятельному решению задач из учебников и пособий, входящих в список основной литературы.

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.**

Microsoft Office (Excel, Power Point), Anaconda, Python, Jupyter Notebook

**12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.**

Компьютерный класс, оборудованный для проведения лекционных и практических занятий средствами оргтехники, персональными компьютерами, объединенными в сеть с выходом в Интернет; установленное лицензионное и свободное программное обеспечение.

**Материально-техническое обеспечение дисциплины**

Дисциплина «*Основы работы с большими данными*» обеспечена:

**1. Для проведения занятий лекционного типа:**

учебной аудиторией, оборудованной учебной мебелью, мультимедийными средствами обучения для демонстрации лекций-презентаций;

**2. Для проведения занятий семинарского типа :**

лабораторией экономико-управленческих дисциплин, оснащенной следующим оборудованием: рабочее место преподавателя: стол, стул; рабочие места обучающихся: столы компьютерные, стулья; персональные компьютеры с установленным ПО, подключенные к сети Интернет; многофункциональное устройство;

**3. Для самостоятельной работы:**

помещением для самостоятельной работы, оснащенным компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде университета.