



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

**ФОНД  
ОЦЕНОЧНЫХ СРЕДСТВ**

ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

**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**


Фонд оценочных средств по дисциплине "Основы работы с большими данными" составлен 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

Фонд оценочных средств по дисциплине " Основы работы с большими данными" одобрен на заседании кафедры прикладной математики от «20» 01 2023г.,

протокол № 5  Зав. кафедрой \_\_\_\_\_ Кадиев Р.М.

на заседании Методической комиссии факультета математики и компьютерных наук от «25» 01 2023 г., протокол № 4

Председатель  Ризаев М.К.

Фонд оценочных средств по дисциплине " Основы работы с большими данными " согласован с учебно-методическим управлением «УО» февраль 2023 г.

Начальник УМУ  Гасангаджиева А.Г.

(подпись)

Рецензент (эксперт):

ФГБОУ ВО ДГУ, ФИиИТ, зав. каф. ИТиБКС  
З.Х. \_\_\_\_\_

(полное наименование организации и должности руководителя)



Ахмедова

# 1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ по дисциплине " Основы работы с большими данными "

## 1.1. Основные сведения о дисциплине

Общая трудоемкость дисциплины составляет 3 зачетных единиц (108 академических часов).

Вид работы	2 семестр	всего
	<b>Общая трудоёмкость</b>	<b>108</b>
<b>Контактная работа:</b>		
Лекции (Л)	14	14
Практические занятия (ПЗ)	14	14
Лабораторные занятия (ЛЗ)	--	--
Консультации		
Промежуточная аттестация (зачет, экзамен)	<b>экзамен</b>	
<b>Самостоятельная работа:</b>	<b>44</b>	<b>44</b>
- подготовка к контрольной работе;	10	10
- самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий);	12	12
- подготовка к практическим занятиям;	10	10
- подготовка к коллоквиумам;	12	12
- подготовка к рубежному контролю)		
<b>Вид итогового контроля (зачет, экзамен, дифференцированный зачет)</b>	<b>экзамен</b>	<b>36</b>

## 1.2. Требования к результатам обучения по дисциплине, формы их контроля и виды оценочных средств

№ п/п	Контролируемые модули, разделы, (темы) дисциплины, их наименование	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	Способ контроля
1	<b>Определение Больших данных. Метаданные</b>	ОПК-1, ОПК-2, ПК-6, ПК-7	Доклад. Практическая работа	Устно Проверка практической работы
2	<b>Параллельные алгоритмы. Программные платформы</b>	ОПК-1, ОПК-2, ПК-6, ПК-7	Устный опрос Практическая работа	Устно Проверка практической работы

## 1.3. Показатели и критерии определения уровня сформированности компетенций

№ п/п	Код компетенции	Уровни сформированности компетенции			
		Недостаточный	Удовлетворительный (достаточный)	Базовый	Повышенный
1	<b>ОПК-1</b>	<p><i>Не знает, как решать</i> осуществлять статистическое наблюдение с использованием стандартных методик и технических средств, включая выборочной совокупности и подготовку статистического инструментария</p> <p><i>Не умеет</i> собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p> <p><i>Не владеет</i> статистическими методами обработки информации, в том числе с применением информационно-коммуникационных технологий.</p>	<p><i>Знает</i> на достаточном уровне осуществлять статистическое наблюдение с использованием стандартных методик и технических средств, включая выборочной совокупности и подготовку статистического инструментария</p> <p><i>Умеет</i> на достаточном уровне собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p> <p><i>Владеет</i> на достаточном уровне статистическими методами обработки информации, в том числе с применением информационно-коммуникационных технологий</p>	<p><i>Знает</i> на хорошем уровне осуществлять статистическое наблюдение с использованием стандартных методик и технических средств, включая выборочной совокупности и подготовку статистического инструментария</p> <p><i>Умеет</i> на хорошем уровне собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p> <p><i>Владеет</i> на хорошем уровне собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p>	<p><i>Знает</i> в совершенстве осуществлять статистическое наблюдение с использованием стандартных методик и технических средств, включая выборочной совокупности и подготовку статистического инструментария</p> <p><i>Умеет</i> в совершенстве собрать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p> <p><i>Владеет</i> в совершенстве собирать исходные данные об объекте исследования и выбрать соответствующий инструментарий для обработки информации</p>
	<b>ОПК-2</b>	<p><i>Не знает</i> основные модели решения функциональных и вычислительных задач,</p> <p><i>Не умеет</i> применять основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач.</p> <p><i>Не владеет</i> основными моделями решения</p>	<p><i>Знает</i> на достаточном уровне основные модели решения функциональных и вычислительных задач, инструментальные средства для решения прикладных задач.</p> <p><i>Умеет</i> на достаточном уровне планировать и проводить аналитические работы, в том числе с применением технологий больших</p>	<p><i>Знает</i> на хорошем уровне основные модели решения функциональных и вычислительных задач.</p> <p><i>Умеет на хорошем уровне:</i> проводить аналитические работы, в том числе с применением технологий больших данных</p> <p><i>Владеет на хорошем уровне</i> основными моделями решения функциональных и вычислительных</p>	<p><i>Знает</i> в совершенстве основные модели решения функциональных и вычислительных задач.</p> <p><i>Умеет в совершенстве</i> проводить аналитические работы, в том числе с применением технологий больших данных.</p> <p><i>Владеет в совершенстве</i></p>

		функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач.	данных. Владеет на достаточном уровне основными моделями решения функциональных и вычислительных задач, инструментальными средствами для решения прикладных задач	задач, инструментальными средствами для решения прикладных задач	основными моделями решения функциональных и вычислительных задач, инструментальным и средствами для решения прикладных задач
3	<b>ПК-6.</b>	Не знает: формировать источники больших данных. Умеет: формировать источники больших данных. Владеет: навыками формирования источников больших данных.	Знает на достаточном уровне формировать источники больших данных. Умеет на достаточном уровне формировать источники больших данных. Владеет на достаточном уровне навыками формирования источников больших данных.	Знает на хорошем уровне формировать источники больших данных. Владеет на хорошем уровне навыками формирования источников больших данных.	Знает в совершенстве формировать источники больших данных. Умеет в совершенстве формировать источники больших данных. Владеет в совершенстве навыками формирования источников больших данных.
4	<b>ПК-7.</b>	Не знает: осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Не умеет: осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Не владеет: навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.	Знает на достаточном уровне осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Умеет на достаточном уровне: осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Владеет на достаточном уровне навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.	Знает на хорошем уровне осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Умеет на хорошем уровне осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Владеет на хорошем уровне навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.	Знает в совершенстве осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Умеет в совершенстве : осуществлять выбор методов и инструментальных средств анализа больших данных для решения профессиональных задач. Владеет в совершенстве : навыками выбора методов и инструментальных средств анализа больших данных для решения профессиональных задач.

## **2. КОНТРОЛЬНЫЕ ЗАДАНИЯ И ИНЫЕ МАТЕРИАЛЫ ОЦЕНКИ знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения дисциплины "Основы работы с большими данными»**

### **Задания для решения. Кейс/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

##### **Вопросы для 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

***Комплект 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.





Оценка «хорошо» выставляется студенту, если он твердо знает материал курса, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения;

Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических задач;

Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями решает практические задачи или не справляется с ними самостоятельно

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.