

*Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Северо-Осетинский государственный университет  
имени Коста Левановича Хетагурова»*



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
«Методы и системы обработки больших данных»**

Направление подготовки 01.03.02 Прикладная математика и информатика

Профиль: Математическое моделирование и вычислительная математика

**Форма обучения – очная**

Владикавказ, 2019

Программа составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 01.03.02 Прикладная математика и информатика, утвержденным приказом Министерства образования и науки Российской Федерации от 10.01.2018 г. № 9, учебным планом подготовки бакалавриата по направлению 01.03.02 Прикладная математика и информатика, профиль: «Математическое моделирование и вычислительная математика», утвержденным Ученым советом ФГБОУ ВО «СОГУ» от 28.05.2019 г. № 10.

Составитель: к.ф.-м.н. Басаева Е.К.

Рабочая программа

обсуждена и утверждена на заседании кафедры прикладной математики (протокол № 8 от 14.03.2019г.);

одобрена советом факультета математики и информационных технологий (протокол № 5 от 29.03.2019 г.).

## 1. Структура и общая трудоемкость дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы. (72 часа).

	Очная Форма обучения
Курс	4
Семестр	7
Лекции	18
Практические занятия	-
Лабораторные занятия	16
Консультации	+
Итого аудиторных занятий	34
Самостоятельная работа	11
Курсовая работа	-
Зачет	27
Экзамен	
Общее количество часов	72

## 2. Цели освоения дисциплины

Цель освоения дисциплины – овладение алгоритмами, парадигмами и инструментами для пакетной и потоковой обработки больших объёмов данных; приобретение навыков проектирования архитектур, применения инструментов разработки для анализа больших объёмов данных.

## 3. Место дисциплины в структуре ОПОП:

Б1.В.ДВ.05.01. Блок 1. Дисциплины (модули). Часть, формируемая участниками образовательных отношений. Дисциплины по выбору.

Для освоения данной дисциплины полезны знания, умения, навыки и компетенции, формируемые предшествующими дисциплинами «Алгебра» «Аналитическая геометрия», «Математический анализ», «Теория вероятностей и математическая статистика», «Алгоритмы и структуры данных», «Основы и методология программирования», «Вычислительная математика», «Введение в анализ данных», «Проектный семинар "Машинное обучение и приложения"».

Знания, навыки и умения, полученные при освоении этой дисциплины, могут быть использованы при изучении дисциплины «Методы и системы обработки больших данных».

## 4. Требования к результатам освоения дисциплины

Компетенции обучающегося, формируемые в результате освоения дисциплины.

УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач;

УК-2 – Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений;

ПК-2 – Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники.

Взаимосвязь планируемых результатов обучения по дисциплине с формируемыми компетенциями ОПОП:

Компетенции		Планируемые результаты обучения, соответствующие формируемым компетенциям ОПОП		
Код	Формулировка			
		Знать:	Уметь	Владеть:
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач/	– методы осуществления поиска и анализа больших данных.	– применять статистические и математические методы для анализа больших объемов информации.	– навыками применения статистических методов для обработки и анализа больших объемов информации.
УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	– типы хранилищ больших объёмов данных и их внутреннее устройство;  – технологии подготовки, хранения, обработки и анализа больших данных.	– решать задачи статистики, задачи поиска и индексации, задачи машинного обучения.	– навыками решения задач машинного обучения, основанных на больших данных.
ПК-2	Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники	– подходы к потоковой и пакетной обработке данных	– диагностировать и решать базовые проблемы инфраструктуры обработки больших данных	– практическими навыками создания моделей, основанных на больших данных, с помощью программ R-Studio и Python.

## 5. Содержание и учебно-методическая карта дисциплины

Таблица 5.1

Номер недели	Наименование тем (вопросов), изучаемых по данной дисциплине	Занятия			Самостоятельная работа студентов		Формы контроля	Баллы		Литература
		л	пр	лаб	Содержание	Часы		min	max	
1–2	<b>Ведение в анализ больших данных. Обзор источников информации.</b> Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных.	2		2			Опрос, лабораторная работа	0	20	[1]–[3], [5], [6]
3–5	<b>Технологии хранения и обработки больших данных.</b> Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.	4		2			Опрос, лабораторная работа			
7, 9, 11, 13	<b>Статистические методы анализа данных.</b> Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).	8					Опрос, доклад, лабораторная работа	0	20	[1]–[3], [5]–[7]
6, 8, 10, 12, 14	<b>Методы обработки и анализа больших данных.</b> Представление исходных данных в программе R-Studio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах Excel и R-Studio: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного			10	Статистическая обработка данных в программах в Python	11		0	30	[2], [4], [7], [8]

	типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы).									
15, 17	<b>Современные программные средства анализа больших объемов информации.</b> Обзор современных популярных программных средства анализа данных: Statistica, SPSS, Excel, R-Studio, Python и др.; их преимущества и недостатки.	4					Опрос	0	10	[2], [4], [8]
16, 18	<b>Визуализация исходной информации и аналитических данных.</b> Возможности графического представления информации в программе R-Studio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.			4			лабораторная работа	0	20	[8]
	<b>ИТОГО</b>	18	0	16		11		<b>0</b>	<b>100</b>	

**Примечания:**

- Все виды учебной работы могут проводиться дистанционно на основании локальных нормативных актов.
- В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по индивидуальной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины может осуществляться через индивидуальные консультации преподавателя очно, в часы консультаций, по электронной почте и с использованием платформ дистанционного обучения.

## 6. Образовательные технологии

При преподавании дисциплины в основном используются традиционные образовательные технологии: лекции, лабораторные занятия и самостоятельная работа студентов. Также при проведении занятий могут быть использованы современные интерактивные и информационно-коммуникационные образовательные технологии такие как:

- интерактивная лекция представляет собой выступление преподавателя перед студенческой аудиторией с применением следующих интерактивных форм обучения: управляемая дискуссия или беседа; демонстрация слайдов или учебных фильмов; мотивационная речь и др.;

- видеоконференция – технология интерактивного взаимодействия двух и более участников образовательного процесса для обмена информацией в реальном режиме времени;

- онлайн-семинар – разновидность веб-конференции, проведение онлайн-встреч или презентаций через Интернет в режиме реального времени;

- творческое задание требует от студента воспроизведения полученной ранее информации в форме, определяемой преподавателем и требующей творческого подхода: подбор материалов по заданной теме; подбор примеров из практики; самостоятельная постановка и решение нетиповых практических задач;

- презентация проекта – слайд-презентации позволяют эффектно и наглядно представить содержание, выделить и проиллюстрировать сообщение.

## 7. Учебно-методическое обеспечение самостоятельной работы

По дисциплине предусмотрены следующие виды самостоятельной работы студентов:

- самостоятельное повторение и изучение теоретического материала;
- подготовка доклада по теме, вынесенной на самостоятельное изучение;
- подготовка к лабораторным занятиям и зачету.

Для обеспечения самостоятельной работы студентов предусмотрены следующие методические материалы (см. разделы 8–9): перечень вопросов для подготовки к зачету, перечень рекомендованной литературы. Учебно-методические материалы необходимые для обеспечения самостоятельной работы студентов размещаются на дистанционной площадке СОГУ в начале каждого модуля.

## 8. Оценочные средства для текущего контроля успеваемости, рубежной аттестации и промежуточной аттестации по итогам освоения дисциплины

*Основными формами текущего контроля* по дисциплине являются устный опрос на практическом занятии, доклад по теме, вынесенной на самостоятельное изучение, выполнение лабораторных работ.

*Формы промежуточной аттестации:* зачет в конце 6 семестра.

Студенты, набравшие в ходе текущего и рубежного контроля 56–100 баллов, автоматически получают «Зачет» или соответствующую шкале экзаменационную оценку. Результирующая оценка рассчитывается по согласно действующей бально-рейтинговой системе.

### 8.1. Формы контроля и критерии оценивания

Этап	Форма контроля	Критерии оценивания (процент от максимального кол-ва баллов)			
		86-100 %	71–85%	56–70%	Менее 56%
		отлично / зачет	хорошо / зачет	удовлетвори-тельно / зачет	неудовлетвори-тельно / незачет
1. Текущий контроль (тах 50 баллов за семнстр)					
		2 балла	1 балл	0,5 балла	0 баллов

	устный опрос (до 10 баллов за семестр)	Тема полностью раскрыта. Превосходное владение материалом. Высокий уровень самостоятельности, логичности, аргументированности. Превосходный стиль изложения.	Тема в основном раскрыта. Хорошее владение материалом. Средний уровень самостоятельности, логичности, аргументированности. Хороший стиль изложения.	Тема частично раскрыта. Удовлетворительное владение материалом. Низкий уровень самостоятельности, логичности, аргументированности. Удовлетворительный стиль изложения.	Тема не раскрыта. Неудовлетворительное владение материалом. Недостаточный уровень самостоятельности, логичности, аргументированности. Неудовлетворительный стиль изложения.
	устный доклад (до 15 баллов за семестр)	5 баллов Тема полностью раскрыта. Превосходное владение материалом. Высокий уровень самостоятельности, логичности, аргументированности. Превосходный стиль изложения.	4 балла Тема в основном раскрыта. Хорошее владение материалом. Средний уровень самостоятельности, логичности, аргументированности. Хороший стиль изложения.	3 балла Тема частично раскрыта. Удовлетворительное владение материалом. Низкий уровень самостоятельности, логичности, аргументированности. Удовлетворительный стиль изложения.	0-2 балла Тема не раскрыта. Неудовлетворительное владение материалом. Недостаточный уровень самостоятельности, логичности, аргументированности. Неудовлетворительный стиль изложения.
	лабораторная работа (до 25 баллов за семестр)	5 баллов Правильно выполнены все задания. Продемонстрирован высокий уровень владения материалом. Проявлены превосходные способности применять знания и умения к выполнению конкретных заданий.	4 балла Правильно выполнена большая часть заданий. Присутствуют незначительные ошибки. Продемонстрирован хороший уровень владения материалом. Проявлены средние способности применять знания и умения к выполнению конкретных заданий.	3 балла Задания выполнены более чем наполовину. Присутствуют серьезные ошибки. Продемонстрирован удовлетворительный уровень владения материалом. Проявлены низкие способности применять знания и умения к выполнению конкретных заданий.	0-2 балла Задания выполнены менее чем наполовину. Продемонстрирован неудовлетворительный уровень владения материалом. Проявлены недостаточные способности применять знания и умения к выполнению конкретных заданий.
	Суммарная оценка текущей работы студента в течение семестра	43–50 баллов Студент активно работает на занятиях, превосходно выполняет все задания преподавателя.	38–42 баллов Студент активно работает на занятиях, хорошо выполняет задания преподавателя.	30–37 баллов Студент недостаточно активно работает на занятиях, удовлетворительно выполняет задания преподавателя.	0–29 баллов Студент недостаточно активно работает на занятиях, неудовлетворительно выполняет задания преподавателя.
<b>2. Итоговый контроль по дисциплине (max 50 баллов)</b>					
		43–50 баллов	36–42 балла	28–35 баллов	0–27 баллов
	Устный экзамен/зачет	Дан полный, развернутый ответ на поставленные вопросы. Ответ формулируется в терминах науки, изложен литературным языком, логичен, доказателен, демонстрирует авторскую позицию студента.	Дан полный ответ на поставленные вопросы, показано умение выделить существенные и несущественные признаки, причинно-следственные связи. Но допущены незначительные ошибки, исправленные студентом с помощью «наводящих» вопросов преподавателя.	Дан недостаточно полный ответ. Студент не способен самостоятельно выделить существенные и несущественные признаки и причинно-следственные связи. Речевое оформление требует поправок, коррекции.	Не получены ответы по базовым вопросам дисциплины или дан неполный ответ и допущены грубые ошибки. Речь неграмотная. Уточняющие вопросы преподавателя не приводят к коррекции ответа студента не только на поставленный вопрос, но и на другие вопросы дисциплины.



## **8.2. Вопросы для подготовки к текущим опросам, лабораторным работам, экзамену.**

1. Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных.
2. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных.
3. Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных.
4. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.
5. Параметрические, непараметрические и номинальные методы анализа данных.
6. Корреляционный анализ.
7. Регрессионный анализ.
8. Дисперсионный анализ.
9. Кластерный анализ.
10. Дискриминантный анализ: модель и общая процедура выполнения.
11. Факторный анализ.
12. Программные средства анализа данных: Statistica, SPSS, Excel, Python; их преимущества и недостатки.
13. Преимущества работа с данными в программе R-Studio.
14. Представление исходных данных в программе R-Studio.
15. Выполнение анализа данных в R-Studio.

## **9. Учебно-методическое и информационное обеспечение дисциплины**

### **а) основная литература:**

1. Каган, Е.С. Прикладной статистический анализ данных: учебное пособие / Е.С. Каган; Кемеровский государственный университет. – Кемерово: Кемеровский государственный университет, 2018. – 235 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=573550>. – Текст: электронный.
2. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel: учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/400278>.
3. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2018. — 121 с. — (Университеты России). — ISBN 978-5-534-09837-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/428765>
4. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python : [16+] / В.М. Волкова, М.А. Семёнова, Е.С. Четвертакова, С.С. Вожов ; Новосибирский государственный технический университет. – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576496>. – Текст : электронный.
5. Колемаев, В.А. Теория вероятностей и математическая статистика: учебник / В.А. Колемаев, В.Н. Калинина. – Москва : Юнити, 2015. – 352 с. : табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=436721>. – Текст: электронный.

### **б) дополнительная литература:**

6. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебник для прикладного бакалавриата / В. Е. Гмурман. — 12-е изд. — Москва : Издательство Юрайт, 2018. — 479 с. — (Бакалавр. Прикладной курс). — ISBN 978-5-534-00211-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/412456>.

7. Александровская, Ю.П. Классификация многомерных данных в экономике: дискриминантный анализ / Ю.П. Александровская ; Казанский национальный исследовательский технологический университет. — Казань : Казанский научно-исследовательский технологический университет (КНИТУ), 2018. — 80 с. : схем., табл., ил. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=500437>. — Текст : электронный.

8. Шипунов, А. Б. Наглядная статистика. Используем R! / А. Б. Шипунов, Е. М. Балдин, П. А. Волкова, А. И. Коробейников, С. А. Назарова, С. В. Петров, В. Г. Суфиянов. - Москва : ДМК Пресс, 2012. - 298 с. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785940748281.html>.

#### **в) Электронные библиотечные системы, профессиональные базы данных и другие интернет-ресурсы**

- ЭБС «Научная электронная библиотека eLibrary.ru» (<http://www.elibrary.ru>).
- ЭБС «Университетская библиотека online» (<http://www.biblioclub.ru>)
- ЭБС «Юрайт» (<http://biblio-online.ru>)
- ЭБС «Консультант студента» ([studentlibrary.ru](http://studentlibrary.ru))
- Федеральная служба государственной статистики. Центральная база статистических данных. <http://www.gks.ru/>

### **10. Материально-техническое обеспечение дисциплины**

Для проведения занятий используются:

– учебные аудитории для проведения лекционных занятий, занятий семинарского типа, текущего контроля, промежуточной аттестации, оборудованные аудиторной мебелью, доской (меловой, маркерной или интерактивной), мультимедийным проектором, экраном, компьютером или ноутбуком с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду СОГУ.

– компьютерный класс (корпус 10, ауд. № 600, 601, 606), оборудованный аудиторной мебелью, доской (меловой или маркерной), компьютерами или ноутбуками с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде СОГУ, мультимедийным проектором, экраном.

*Лицензионное программное обеспечение:*

1. Windows 10 Pro for Workstations, (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
2. Office Standard 2016 (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
3. Система поиска текстовых заимствований «Антиплагиат ВУЗ»;

*Перечень ПО в свободном доступе:* Kaspersky Free; WinRar; Google Chrome; Yandex Browser; Opera Browser; Acrobat Reader; LaTeX; системы компьютерной алгебры SciLab и Maxima, R, Python, Anaconda.

Помещение для самостоятельной работы студентов: Зал электронных ресурсов Научной библиотеки СОГУ (корпус 6, кабинет № 1.8), укомплектован специализированной мебелью (рабочие места студентов), необходимыми техническими средствами обучения: компьютеры, принтер, возможность подключения к сети «Интернет», доступ в электронную информационно-образовательную среду СОГУ.

## **11. Лист обновления/актуализации**

### **1. Рабочая программа**

пересмотрена и актуализирована на заседании кафедры прикладной математики, протокол № 7 от 19.03.2019г.;

одобрена на заседании совета факультета математики и информационных технологий, протокол № 5 от 27.03.2020 г.