

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Кабардино-Балкарский государственный университет  
им. Х.М. Бербекова» (КБГУ)

ИНСТИТУТ ИСККУСТВЕННОГО ИНТЕЛЛЕКТА И ЦИФРОВЫХ ТЕХНОЛОГИЙ  
КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ ИНФОРМАЦИОННОЙ  
БЕЗОПАСНОСТИ

СОГЛАСОВАНО

Руководитель образовательной  
программы \_\_\_\_\_ Д.А. Крымшожалова  
« \_\_\_\_ » \_\_\_\_\_ 2024 г.

УТВЕРЖДАЮ

Директор института ИИИЦТ  
\_\_\_\_\_ З.В. Шомахов  
« \_\_\_\_ » \_\_\_\_\_ 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

«Методы обработки больших данных»

по направлению подготовки  
**09.03.01 – Информатика и вычислительная техника**

Профиль:  
Интеллектуальные системы обработки информации и  
управления

Квалификация (степень) выпускника  
Бакалавр

Форма обучения  
очная форма

Нальчик 2024

Рабочая программа дисциплины «Методы обработки больших данных» /сост. Хаширова Т.Ю. – Нальчик: КБГУ, 2024. 23 с.

Рабочая программа предназначена для преподавания дисциплины «Методы обработки больших данных» вариативной части студентам очной формы обучения, по направлению подготовки 09.03.01 – Информатика и вычислительная техника, для профиля Интеллектуальные системы обработки информации и управления, в 5 семестре, 3 курса.

Рабочая программа дисциплины составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки Российской Федерации от «19» сентября 2017 г. №929 (зарегистрировано в Минюсте России 10 октября 2017 г. № 48489).

## СОДЕРЖАНИЕ

1. Цели и задачи изучения дисциплины.....	4
2. Место дисциплины в структуре ОПОП ВО.....	4
3. Требования к результатам освоения дисциплины .....	5
4. Содержание и структура дисциплины .....	5
5. Оценочные материалы для текущего и рубежного контроля успеваемости и промежуточной аттестации.....	7
6. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности .....	19
7. Учебно-методическое и информационное обеспечение дисциплины.....	19
8. Материально-техническое обеспечение дисциплины .....	24
9. Лист переутверждения рабочей программы дисциплины .....	26
Приложение .....	27

## 1. ЦЕЛИ И ЗАДАЧИ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

Курс имеет своей целью: формирование у студентов профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных. Данная цель соотносится с целью образовательной программой в частности с технологий разработки специализированных программных систем, отвечающих за обработку больших данных. Изучение данной дисциплины готовит выпускника к выполнению следующих профессиональных задач:

- Постановка задачи анализа данных.
- Предварительная обработка данных.
- Визуализация данных.
- Разработка, реализация и применение методов интеллектуального анализа данных к большим массивам данных.
- Представление результатов работы.

В результате освоения дисциплины студент должен овладеть компетенциями ПКС-2; ПКС-3.

ПКС-2 Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности;

ПКС-3 Способен разрабатывать графический дизайн интерфейса, проектировать пользовательские интерфейсы по готовому образцу или концепции интерфейса.

Основные задачи освоения дисциплины:

Студент должен знать методы анализа и хранения больших объемов данных, этапы жизненного цикла обработки больших данных, языки, наиболее приспособленные для обработки и аналитики больших данных, способы организации хранения и доступа к большим данным; уметь выполнять элементы анализа данных и интерпретировать результаты, различать характеристики SQL и NoSql БД, формулировать алгоритмы в парадигме MapReduce, выбрать подходящий инструмент анализа больших данных, выбрать подходящую технологию хранения больших данных.; владеть математическими методами анализа данных, языками и компьютерными методами обработки.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина Б1.В.ДВ.01.01 «Методы обработки больших данных» относится к части, формируемой участниками образовательных отношений, дисциплина по выбору), предназначена для преподавания студентам очной формы обучения на 3 курсе в 5 семестре..

Основные знания, необходимые для изучения дисциплины: информатика, программирование, численные методы и математические дисциплины, предусмотренные учебным планом. На изучение курса отводится 108 часов (3 з.е.), из них лекционных – 17, лабораторных – 34, самостоятельная работа студента – 48 часов, заканчивается зачетом.

Дисциплина позволит расширить теоретическую подготовку бакалавра, углубить знание прикладных вопросов, связанных с использованием современных систем программирования и программных продуктов для решения задач профессиональной деятельности. Знания, полученные студентами в рамках освоения дисциплины, востребованы организациями и специалистами, применяющими в своей деятельности математические методы моделирования, обработку экспериментальных данных.

В результате освоения дисциплины обучающийся сможет частично продемонстрировать следующие обобщенные трудовые функции (ОТФ):

- Разработка требований и проектирование программного обеспечения (профессиональный стандарт 06.001 – «Программист», код D, уровень квалификации – 6).

- Концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности (профессиональный стандарт 06.022 – «Системный аналитик», код С, уровень квалификации – 6).

### 3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по направлению подготовки 09.03.01 «Информатика и вычислительная техника»:

ПКС-2 Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности;

ПКС-3 Способен разрабатывать графический дизайн интерфейса, проектировать пользовательские интерфейсы по готовому образцу или концепции интерфейса. **Коды и**

В результате изучения дисциплины студенты должны:

знать

- суть основных подходов к обработке экспериментальной информации;
- смысл и назначение визуальных (графических) методов представления информации;
- основные статистические понятия и методы исследования;
- современные системы, применяемые для различных видов анализа пространственных данных;

уметь

- проводить различного вида анализ исходных данных;
- планировать эксперимент, т.е. выбирать план эксперимента, удовлетворяющего заданным требованиям, совокупность действий, направленных на разработку стратегии экспериментирования (от получения априорной информации до получения работоспособной математической модели или определения оптимальных условий). Это целенаправленное управление экспериментом, реализуемое в условиях неполного знания механизма изучаемого явления;
- в процессе проведения необходимых измерений и, последующей обработки данных формализовать результат в виде математической модели;
- оценивать возникающие погрешности, содержащейся в исходных данных.

### 4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

В таблице 1 приводится описание содержания дисциплины, структурированное по разделам, с указанием по каждому разделу формы текущего контроля: защита лабораторной работы (ЛР), коллоквиум (К), рубежный контроль (РК), тестирование (Т).

Таблица 1

#### Содержание разделов дисциплины «Методы обработки больших данных»

№	Наименование раздела	Содержание раздела	Код контролируемой компетенции (или ее части)	Форма текущего контроля
1.	<b>Введение в большие данные.</b> Понятие DataMinig. Прикладные инструменты для работы с Big Data. Технология MapReduce. Hadoop.	Основы теории подобия (точное подобие, Основные термины и определения; классификация методов планирования эксперимента; научный эксперимент.	ПКС-2 ПКС-3	К, Т

2.	<b>Технологии анализа данных:</b> Жизненный цикл анализа больших данных, стандарты. Когнитивный анализ данных. Визуализация больших данных.	Основы теории планирования эксперимента. Последовательность проведения исследований	ПКС-2 ПКС-3	К, РК, Т
3.	<b>Технологии хранения больших данных.</b> Распределенные хранилища, NoSql хранилища, классификация и примеры.	Однофакторный дисперсионный анализ; двухфакторный дисперсионный анализ;	ПКС-2 ПКС-3	К, РК, Т

Общая трудоемкость дисциплины составляет **3 зачетные единицы (108 часов)**

Таблица 2 Структура дисциплины «Методы обработки больших данных»

Вид работы	Трудоемкость, часы	Всего
	5 семестр	
<b>Общая трудоемкость (в зачетных единицах)</b>	3	3
<b>Контактная работа (в часах):</b>	<b>51</b>	<b>51</b>
<i>Лекции (Л)</i>	<i>17</i>	<i>17</i>
<i>Практические занятия (ПЗ)</i>	-	-
<i>Семинарские занятия (СЗ)</i>	-	-
<i>Лабораторные работы (ЛР)</i>	<i>34</i>	<i>34</i>
<b>Самостоятельная работа (в часах):</b>	<b>48</b>	<b>48</b>
Курсовой проект (КП), Курсовая работа (КР)	-	-
Расчетно-графическое задание (РГЗ)	-	-
Реферат (Р)	-	-
Эссе (Э)	-	-
Самостоятельное изучение разделов	48	48
Контрольная работа (К)	-	-
Подготовка и прохождение промежуточной аттестации	9	9
<b>Вид промежуточной аттестации</b>	<b>зачет</b>	<b>зачет</b>

Таблица 3. Лекционные занятия

№	Содержание раздела
1.	Предпосылки формирования тренда больших данных, Основные вызовы больших данных, Определение термина "большие данные", Базовое представление о Map Reduce и Hadoop, Представление о работе аналитика, Инструменты для обработки больших данных, знакомство с языками и прикладными, пакетами для обработки больших данных, Рассмотрение общей концепции и синтаксиса языка R (примеры).
2.	Аналитика больших данных.
3.	Когнитивный анализ данных
4.	Методы DATA MINING
5.	Хранилища данных. Регрессия
6.	Распределенные базы данных NoSQL.

Таблица 4. Практические занятия - не предусмотрены.

Таблица 5. Лабораторные работы

№	Наименование тем лабораторных работ
---	-------------------------------------

1.	Ознакомление с синтаксисом языка R для анализа данных.(4ч)
2.	Способы подготовки и отображения данных в R (4ч). Возможности ввода/вывода.
3.	Решение задач на больших графах (2ч).
4.	Способы анализа данных в R. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных(2ч).
5.	Работа с диаграммами и графиками в R (2ч).
6.	Проверка статистических гипотез (4ч)
7.	Корреляционный анализ и регрессионный анализ данных (2ч)

Таблица 6. Самостоятельное изучение разделов дисциплины

№ раз-дела	Наименование тем
1.	Работа с диаграммами и графиками в R (2ч).
2.	Проверка статистических гипотез (4ч)
3.	Корреляционный анализ и регрессионный анализ данных (2ч)
4.	Решение задач Data Mining.
5.	Задачи классификации, кластеризации: деревья решений, RandomForest, k-means. (4ч)
6.	Изучение принципов работы распределенных баз данных
7.	Развертывание локального кластера Hadoop. Подсчёт слов в тексте, с помощью MapReduce. (4ч)

## 5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО И РУБЕЖНОГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Формы контроля текущих, рубежных и промежуточных знаний студентов по дисциплине определяются в соответствии с учебным планом образовательной программы и в соответствии с действующим Положением о балльно-рейтинговой системе оценки успеваемости студентов КБГУ.

От обучающихся требуется посещение занятий, выполнение лабораторных работ, знакомство с рекомендованной литературой.

При аттестации обучающихся оценивается качество работы на занятиях (умение вести дискуссию, способность четко и ёмко формулировать свои мысли), уровень подготовки к самостоятельной деятельности, качество выполнения заданий (презентаций, докладов, выполнение лабораторных работ и др.).

Конечными результатами освоения программы дисциплины являются сформированные когнитивные дескрипторы «знать», «уметь», «владеть», расписанные по отдельным компетенциям. Формирование этих дескрипторов происходит в течение всего семестра по этапам в рамках различного вида занятий и самостоятельной работы.

В ходе изучения дисциплины предусматриваются текущий, рубежный контроль и промежуточная аттестация. Формы контроля текущих, рубежных и промежуточных знаний

студентов по дисциплине определяются в соответствии с учебным планом образовательной программы и в соответствии с действующим Положением о балльно-рейтинговой системе оценки успеваемости студентов КБГУ.

От обучающихся требуется посещение занятий, выполнение лабораторных работ, знакомство с рекомендованной литературой.

При аттестации обучающихся оценивается качество работы на занятиях (умение вести дискуссию, способность четко и ёмко формулировать свои мысли), уровень подготовки к самостоятельной деятельности, качество выполнения заданий (презентаций, докладов, выполнение лабораторных работ и др.).

### **5.1. Оценочные материалы для текущего контроля.**

Цель текущего контроля – оценка результатов работы в семестре и обеспечение своевременной обратной связи, для коррекции обучения, активизации самостоятельной работы обучающегося. Объектом текущего контроля являются конкретизированные результаты обучения (учебные достижения) по дисциплине

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины «Методы обработки больших данных», оценка качества подготовки на основании выполненных заданий ведется преподавателем (с обсуждением результатов), баллы начисляются в зависимости от сложности задания.

#### **Критерии формирования оценок (оценивания) устного опроса**

Устный опрос является одним из основных способов учёта знаний обучающегося по дисциплине. Развёрнутый ответ должен представлять собой связное, логически последовательное сообщение на заданную тему, показывать его умение применять определения.

В результате устного опроса знания, обучающегося оцениваются по следующей шкале:

<b>3 балла</b>	<b>2 балла</b>	<b>1 балл</b>	<b>0 баллов</b>
<p>ставится, если обучающийся:</p> <p>1) полно излагает изученный материал, даёт правильное определенное экономических понятий;</p> <p>2) обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только по учебнику, но и самостоятельно составленные;</p> <p>3) излагает материал последовательно и правильно с точки зрения норм литературного языка.</p>	<p>ставится, если обучающийся даёт ответ, удовлетворяющий тем же требованиям, что и для балла «1», но допускает 1-2 ошибки, которые сам же исправляет, и 1-2 недочёта в последовательности и языковом оформлении излагаемого.</p>	<p>ставится, если обучающийся обнаруживает знание и понимание основных положений данной темы, но:</p> <p>1) излагает материал неполно и допускает неточности в определении понятий;</p> <p>2) не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры;</p> <p>3) излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.</p>	<p>ставится, если обучающийся обнаруживает незнание большей части соответствующего раздела изучаемого материала, допускает ошибки в формулировке.</p>

Баллы «1», «2», «3» могут ставиться не только за единовременный ответ, но и за расчленившийся во времени, т.е. за сумму ответов, данных на протяжении занятия. начисляются в зависимости от сложности задания.

## 5.2 Оценочные материалы для самостоятельной работы обучающегося

Рабочая программа предусматривает проведение лекционных, лабораторных занятий, а также самостоятельную работу обучающихся. В ФГБОУ ВО «Кабардино-Балкарский государственный университет» действует балльно-рейтинговая система оценки учебных достижений обучающихся по образовательным программам, реализуемым на основании федеральных государственных образовательных стандартов. Балльно-рейтинговая система оценки знаний является одной из составляющих системы управления качеством образовательной деятельности в университете.

Таблица 7

Распределение баллов в соответствии с действующим Положением о балльно-рейтинговой системе

№ рейтинговой точки	Коллоквиум	Лаб.практикум	Посещаемость	Тестирование	Итого
1	7	8	3	5	23
2	7	8	3	5	23
3	7	8	4	5	24

Таблица 8

### Критерии оценки

Ви д ме ро пр ият ия	К ри те ри и оц ен ки	Баллы
Ко лло кви ум (ус тн ый оп рос по тем е)	- яс но ст ь, че тк ос ть и до ка за те ль но ст ь из ло ж ен	0-21 балл

	ия от ве то в на во пр ос ы; - вл ад ен ие сп ец иа ль н ы м и те р м и на м и; - си ст е м но ст ь зн ан и й по те м ат ик е	
Ла бор ато рн	- по н и	0 - 4 балла

ое зан яти е	м ан ие це ли и за да ч ра бо т ы - в ы по лн ен ие за да н и й и об ра бо тк а ре зу ль та то в - от че т и за щ ит а ла бо ра то рн ой	
-----------------------	---	--

	ра бо т ы	
Ко мп ью тер ное те- сти ро- ва- ние по раз де- ла м дис ци пл ин ы	Ре зу ль та т ы те ст ир ов ан ия (К ол ич ес тв о ба лл ов ф, ф - до ля пр ав ил ьн о от ве че н н ы х те ст ов по те м е).	0-15 баллов
По	П	0-10 баллов

се ще ние зан яти й	ри бо ле е 3 пр оп ус ка х бе з ув а ж ит ел ьн ой пр ич и н ы за ня ти й ан ну ли ру ю тс я ба лл ы	
За- чет	яс но ст ь, че тк ос ть и до ка за те ль	0-30 баллов

НО СТ Ь ИЗ ЛО Ж ЕН ИЯ ОТ ВЕ ТО В НА ВО ПРО СЫ; - ВЛА Д ЕН ИЕ СП ЕЦ ИА ЛЬ Н Ы М И ТЕ Р М И НА М И; - СИ СТ Е М НО СТ Ь ЗНА НИ Й ПО ТЕ М	
---	--

ат ик е ди сц и пл и н ы в це ло м	
Итоговая я оценка	0-100 баллов

### 5.2.1. Образцы заданий промежуточного и текущего контроля (контролируемые компетенции ПКС-2; ПКС-3)

1. Подобрать данные для таблицы, приведенной ниже и проанализировать их взаимное влияние, отобразить корреляцию:
  - a. Роста ВВП на прирост населения
  - b. Прироста населения на динамику безработицы
  - c. Прирост людей с высшим образованием на рост промышленного производства
  - d. Прирост людей с высшим образованием на развитие науки
  - e. Прирост людей с высшим образованием на динамику доходов на душу населения
  - f. Динамику безработицы на динамику преступности
  - g. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.
  - h. С помощью функции `predict()` (см. лекции и `help()`) постройте прогноз по столбцу, соответствующему варианту.

Годы	Численность населения	Рост ВВП	Динамика безработицы	Динамика промышленного производства	Прирост людей, получивших высшее образование	Развитие науки (высокотехнологичных отраслей)	Динамика доходов на душу населения	Динамика преступности
	1	2	3	4	5	6	7	8
01.01.1990								
...								
01.01.2015								

Отчет по выполнению РГЗ должен содержать:

2. постановку задачи;
3. сформированный набор данных;
4. тексты скриптов на языке R;

5. результаты тестов на проверку гипотез о корреляции, оценка регрессии, вычисление корреляции в текстовом и графическом виде.
6. ясное и подробное пояснение каждого результата, словесную трактовку графиков;
7. выводы;
8. список использованной литературы.

### **5.2.2. Образцы тестовых заданий (контролируемые компетенции ПКС-2; ПКС-3)**

Тестирование обучающихся проводится в онлайн-режиме согласно расписанию, в ЭИОС open.kbsu.ru

#### **2. На какие классы делятся планы по соотношению между количеством оцениваемых неизвестных параметров модели и количеством точек плана эксперимента?**

- ненасыщенные, насыщенные, сверхнасыщенные;
- насыщенные, сверхнасыщенные;
- ненасыщенные, сверхнасыщенные
- ненасыщенные, насыщенные

#### **3. В чем заключается задача оптимизации в теории планирования экспериментов?**

- заключается в нахождение средней функции отклика в области допустимых значений параметров;
- заключается в нахождение полинома функции отклика в области допустимых значений параметров;
- заключается в нахождение бинома функции отклика в области допустимых значений параметров;
- заключается в нахождение экстремума функции отклика в области допустимых значений параметров;

#### **4. На основе какого метода определяются оценки коэффициентов полинома (функции) отклика?**

- равносторонних треугольников;
- наименьших квадратов;
- наибольших трапеций;
- равных трапеций.

#### **5. Что является целью обработки экспериментальных данных?**

- равносторонних треугольников;
- наименьших квадратов;
- наибольших трапеций;
- равных трапеций.

#### **6. Перечислите источники экспериментальных данных.**

- графическая форма, аудиоданные;
- символьная форма, аудиоданные;
- символьная форма, графическая форма;
- символьная форма, графическая форма, аудиоданные.

#### **7. В каких формах могут быть представлены экспериментальные данные?**

- графическая форма, аудиоданные;
- символьная форма, аудиоданные;
- символьная форма, графическая форма;
- символьная форма, графическая форма, аудиоданные.

**8. Что является источниками ошибок в процедурах регистрации и обработки данных?**

- фонетические ошибки, методические ошибки;
- инструментальные ошибки, морфологические ошибки;
- инструментальные ошибки, методические ошибки;
- синтаксические ошибки, методические ошибки

**9. Какими свойствами характеризуется качество оценки параметров распределения?**

- несостоятельность, несмещенность, эффективность, достаточность;
- состоятельность, несмещенность, эффективность, достаточность
- состоятельность, несмещенность, неэффективность, достаточность;
- состоятельность, несмещенность, эффективность, недостаточность.

**10. В чем отличия между корреляционным и регрессионным анализом?**

- Требуется: определить оценки коэффициентов корреляции для всех или только для заданных пар параметров и оценить их значимость; установить количественную взаимосвязь между показателем и факторами.
- Требуется: определить оценки коэффициентов корреляции только для заданных пар параметров и оценить их значимость; установить количественную взаимосвязь между показателем и факторами.
- Требуется: определить оценки коэффициентов корреляции для всех пар параметров и оценить их значимость; установить количественную взаимосвязь между показателями.
- Требуется: определить оценки коэффициентов корреляции для всех или только для заданных пар параметров и оценить их значимость; установить количественную взаимосвязь между факторами.

**Критерии формирования оценок по тестовым заданиям**

Выполнение тестирования оценивается согласно проценту правильных ответов. Максимально возможное количество баллов за тестирование – 5.

**Оценочные материалы для промежуточной аттестации**

Целью промежуточных аттестаций по дисциплине является оценка качества освоения дисциплины обучающимися.

Промежуточная аттестация предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины. Осуществляется в конце семестра и представляет собой итоговую оценку знаний по дисциплине в виде проведения зачета в 7 семестре.

Промежуточная аттестация может проводиться в устной, письменной форме, и в форме тестирования. На промежуточную аттестацию отводится до 30 баллов.

- В
- о 1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.
- р 2. Дать краткую сравнительную характеристику инструментария ПО для анализа данных.
- с 3. Охарактеризовать конструкции языка R Перечислить типы языка R, привести примеры.
- , 4. Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика. Отличия BI от Data Science.
- в 5. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать

н  
о  
с  
и  
м  
ы

степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.

6. Что такое Data Mining? Основные задачи и методы Data Mining. Этапы интеллектуального анализа данных. Методы интеллектуального анализа данных.

7. Что такое ИИ? Декатлон?

8. Роль гипотез в процессе познания. Какие факторы используются для уточнения гипотез?

9. Основные понятия статистики и дескриптивный анализ:

10. Шкалы измерений. Генеральная совокупность и выборка. Нормальное распределение. Уровень статистической достоверности.

11. Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа.

12. Пояснить термин "Линейная регрессия". Привести примеры использования

регрессионного анализа.

13. Классификация и кластеризация – суть и назначение. Метрики. Постановка задачи кластеризации. Методы кластеризации на графах. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.

14. Парадигма Map Reduce. Описать принцип работы. Нарисовать схему. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.

15. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры и инструменты для визуализации.

16. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.

17. OLAP и OLTP системы. Разница.

18. Репликация и шардинг.

19. Требования ACID. CAP-теорема, BASE архитектура

20. NoSql. Классификация NoSql хранилищ. Их особенности. Примеры распределенных хранилищ.

### **Критерии формирования оценок по промежуточной аттестации**

«Отлично» получают обучающиеся, которые свободно ориентируются в материале и отвечают без затруднений. Обучающийся способен к выполнению сложных заданий, постановке целей и выборе путей их реализации. Работа выполнена полностью без ошибок, решено 100% задач;

«Хорошо» получают обучающиеся, которые относительно полно ориентируются в материале, отвечают без затруднений, допускают незначительное количество ошибок. Обучающийся способен к выполнению сложных заданий. Работа выполнена полностью, но имеются не более одной негрубой ошибки и одного недочета, не более трех недочетов. Допускаются незначительные неточности при решении задач, решено 70% задач;

«Удовлетворительно» получают обучающиеся, у которых недостаточно высок уровень владения материалом. В процессе ответа на экзамене допускаются ошибки и затруднения при изложении материала. Обучающийся правильно выполнил не менее 2/3 всей работы или допустил не более одной грубой ошибки и двух недочетов, не более одной грубой и одной негрубой ошибки, не более трех негрубых ошибок, одной негрубой. Обучающийся затрудняется с правильной оценкой предложенной задачи, дает неполный ответ, решено 55% задач;

«Неудовлетворительно» получают обучающиеся, которые допускают значительные ошибки. Обучающийся имеет лишь начальную степень ориентации в материале. В работе число ошибок и недочетов превысило норму для оценки 3 или правильно выполнено менее 2/3 всей работы. Обучающийся дает неверную оценку ситуации, решено менее 50% задач.

## 6. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРЫ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ

Максимальная сумма (100 баллов), набираемая студентом по дисциплине включает две составляющие:

- первая составляющая – оценка регулярности, своевременности и качества выполнения студентом учебной работы по изучению дисциплины в течение периода изучения дисциплины (семестра, или нескольких семестров) (сумма – не более 70 баллов). Баллы, характеризующие успеваемость студента по дисциплине, набираются им в течение всего периода обучения за изучение отдельных тем и выполнение отдельных видов работ.
- вторая составляющая – оценка знаний студента по результатам промежуточной аттестации (не более 30 –баллов).

Критерием оценки уровня сформированности компетенций в рамках учебной дисциплины «Методы обработки больших данных» в 5 семестре является зачет.

Целью промежуточных аттестаций по дисциплине является оценка качества освоения дисциплины обучающимися.

Таблица 9

Результаты освоения формирования, подлежащие проверке

Компетенция по ФГОС	Основные показатели освоения (показатели достижения результата)	Код показателя освоения
ПКС-2. Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности	<b>ИД-1пкс-2 Знать:</b> основные виды диагностических данных и способы их представления, и технологию разработки алгоритмов и программ	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы (раздел 5)
	<b>ИД-2пкс-2 Уметь:</b> писать программный код процедур проверки работоспособности программного обеспечения на выбранном языке программирования.	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы
	<b>ИД-3пкс-2 Владеть:</b> навыками разработки процедуры сбора диагностических данных и навыками работы с типовыми программными продуктами.	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы (раздел 5)
ПКС-3. Способен разрабатывать графический дизайн интерфейса, проектировать пользовательские интерфейсы по готовому образцу или концепции интерфейса	<b>ИД-1пкс-3 Знать:</b> методики использования программных средств для решения практических задач.	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы (раздел 5)
	<b>ИД-2пкс-3 Уметь:</b> использовать программные средства для решения практических задач.	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы
	<b>ИД-3пкс-3 Владеть:</b> навыками использования программных средств для решения практических задач.	Типовые оценочные материалы для устного опроса, типовые тестовые задания, лабораторные работы (раздел 5)

## 7.УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Нормативно-правовая база

1. ГОСТ Р 57412-2017 Компьютерные модели в процессах разработки, производства и

эксплуатации изделий. Общие положения.

2. I  
S  
O

ИЕС 25020:2007 «Разработка программного обеспечения. Требования к качеству и оценка качества программного продукта. Измерительная эталонная модель и руководство».

## **7.2. Основная литература**

1. Ковель А.А. Инженерные аспекты математического планирования эксперимента: монография / Ковель А.А. — Железногорск: Сибирская пожарно-спасательная академия ГПС МЧС России, 2017. — 117 с. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <https://www.iprbookshop.ru/66909.html>.
2. Любимцева О.Л. Блочное планирование эксперимента и анализ данных: учебное пособие / Любимцева О.Л. — Нижний Новгород: Нижегородский государственный архитектурно-строительный университет, ЭБС АСВ, 2018. — 30 с. — ISBN 978-5-528-00276-7. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <https://www.iprbookshop.ru/80885.html> (дата обращения: 25.12.2021).
3. Юдин Ю.В. Организация и математическое планирование эксперимента: учебное пособие / Юдин Ю.В., Майсурадзе М.В., Водолазский Ф.В. — Екатеринбург: Издательство Уральского университета, 2018. — 124 с. — ISBN 978-5-7996-2486-6. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <https://www.iprbookshop.ru/106473.html> (дата обращения: 25.12.2021).

## **7.3. Дополнительная литература**

1. Адлер Ю.П. Планирование эксперимента. — М.: Знание, 1978.
2. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука. — 1976. — 279 с.
3. Асатурян В.И. Теория планирования эксперимента. — М.: Радио и связь, 1983. — 248 с.
4. Бродский В.З. Введение в факторное планирование эксперимента. — М.: Наука, 1976. — 223 с.
5. Успенский А.Б., Федоров В.В. Вычислительные аспекты метода наименьших квадратов при анализе и планировании регрессионных экспериментов. — М.: Изд-во МГУ, 1975. — 167 с.
6. Федоров В.В. Теория оптимального планирования эксперимента. — М.: Наука, 1971. — 312 с.
7. Юдин М.И. Планирование эксперимента и обработка его результатов: Монография. — Краснодар: КГАУ, 2004. — 239 с.
8. Хаширова Т.Ю., Эдгулова Е.К. Обработка экспериментальных данных. Лабораторный практикум. — Нальчик: КБГУ, 2010 — 18 с.

## **7.4. Интернет-ресурсы**

1. <http://www.diss.rsl.ru>
2. <http://www.scopus.com>
3. <http://elibrary.ru>
4. <http://ipbookshop.ru>
5. <http://нэб.пф>
6. <http://lib.kbsu.ru>

## **7.5. Периодические издания**

1. Вестник Воронежского государственного университета. Серия: Системный анализ и

- информационные технологии.
2. Моделирование и анализ информационных систем.
  3. Информатика и системы управления
  4. Информационное общество

## **7.6. Современные профессиональные базы данных**

1. База данных Science Index (РИНЦ) <http://elibrary.ru>
2. Национальная электронная библиотека РГБ <https://нэб.рф>
3. Крупнейшая единая база данных, содержащая аннотации и информацию о цитируемости рецензируемой научной литературы, со встроенными инструментами отслеживания, анализа и визуализации данных. [www.scopus.com](http://www.scopus.com)
4. Самая полная математическая база данных, охватывающая материалы с конца 19 века. zbMath содержит документы, журналы и книги по математике, статистике, информатике, а также машиностроению, физике, естественным наукам и др. [www.zbmath.org](http://www.zbmath.org) (доступ открытый)

## **7.7. Методические указания по организации аудиторных занятий**

В систему средств обучения дисциплины «Методы обработки больших данных» входят учебники, учебные пособия, методические указания, программное и компьютерное обеспечение, образующие единую комплексную среду, позволяющую достигать поставленных целей обучения.

Организационной формой проведения аудиторных занятий по дисциплине являются лекции, практические и лабораторные занятия.

Основная дидактическая цель лекции – формирование ориентировочной основы для последующего усвоения студентами учебными материала.

Педагогическая эффективность лекций по курсу «Методы обработки больших данных» усиливается использованием специально разработанных демонстрационных программ и компьютерных моделей.

Лабораторные занятия проводятся в компьютерных классах института информатики, электроники и компьютерных технологий. Студенты в течение семестра разбиваются на творческие группы и работают над проектом. Студенты учатся работать в коллективе.

В соответствии с учебным планом на каждую тему выделено необходимое количество часов практических занятий, которые проводятся в соответствии с вопросами, рекомендованными к изучению по определенным темам. Обучающиеся должны регулярно готовиться к лабораторным занятиям и участвовать в обсуждении вопросов. При подготовке к занятиям следует руководствоваться конспектом лекций и рекомендованной литературой. Тематический план дисциплины, учебно-методические материалы, а также список рекомендованной литературы приведены в рабочей программе

### **Методические рекомендации при работе над конспектом во время проведения лекции**

В процессе лекционных занятий целесообразно конспектировать учебный материал. Для этого используются общие и утвердившиеся в практике правила, и приемы конспектирования лекций.

Конспектирование лекций ведется в специально отведенной для этого тетради, каждый лист которой должен иметь поля, на которых делаются пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений.

Целесообразно записывать тему и план лекций, рекомендуемую литературу к теме. Записи разделов лекции должны иметь заголовки, подзаголовки, красные строки. Для выделения разделов, выводов, определений, основных идей можно использовать цветные карандаши и фломастеры.

Названные в лекции ссылки на первоисточники надо пометить на полях, чтобы при самостоятельной работе найти и вписать их. В конспекте дословно записываются определения понятий, категорий и законов. Остальное должно быть записано своими словами.

Каждому обучающемуся необходимо выработать и использовать допустимые сокращения наиболее распространенных терминов и понятий.

### **Методические рекомендации по подготовке к лабораторным занятиям**

Лабораторные занятия – составная часть учебного процесса, групповая форма занятий при активном участии обучающихся. Лабораторные занятия способствуют углубленному изучению наиболее сложных проблем науки и служат основной формой подведения итогов самостоятельной работы обучающихся. Целью лабораторных занятий является углубление и закрепление теоретических знаний, полученных обучающимися на лекциях и в процессе самостоятельного изучения учебного материала, а, следовательно, формирование у них определенных умений и навыков.

В ходе подготовки к лабораторному занятию необходимо прочитать конспект лекции, изучить основную литературу, ознакомиться с дополнительной литературой, выполнить выданные преподавателем задания. При этом учесть рекомендации преподавателя и требования программы. Дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы.

Желательно при подготовке к лабораторным занятиям по дисциплине одновременно использовать несколько источников, раскрывающих заданные вопросы.

### **Методические указания по организации самостоятельной работы студентов**

Самостоятельная работа при прохождении дисциплины «Методы обработки больших данных» должна занимать одно из ведущих мест в учебной деятельности студентов. Она должна быть осознана студентами как свободная по выбору, внутренне мотивированная деятельность. Наличие самостоятельной работы студентов является одним из важнейших средств формирования способностей самостоятельно добывать, перерабатывать и практически применять знания. В результате происходит ограничение объясняющей функции преподавателя, переход от описательного объяснения к доказательному, формирование творческого мышления. Самостоятельная работа предполагает осознание цели своей деятельности, принятие учебной задачи, придание ей личного смысла, самоорганизацию в распределении учебных действий во времени, самоконтроль в их выполнении и др.

Организация контролируемой самостоятельной работы студентов. Самостоятельная работа студентов (СРС) при прохождении дисциплины «Методы обработки больших данных» имеет целью научить студента осмысленно и самостоятельно разрабатывать надежные и наглядные модели, заложить основы самоорганизации и самовоспитания с тем, чтобы привить умение в дальнейшем непрерывно повышать свою квалификацию. Должны сочетаться два основных направления организации самостоятельной работы студентов:

- самостоятельная работа в процессе аудиторных занятий, опирающаяся на использование методик и форм организации аудиторных занятий, способных обеспечить высокий уровень самостоятельности студентов и улучшение качества подготовки;
- самостоятельная работа во внеаудиторное время, основным принципом организации которой является перевод всех студентов на индивидуальную работу с переходом от формального выполнения определенных заданий при пассивной роли студента к познавательной активности с формированием собственного мнения при решении поставленных проблемных вопросов и задач.

Для самостоятельной работы с дидактическими материалами по имитационному моделированию студентам предлагаются подробные описания заданий в бумажном и электронном виде.

Помимо описаний заданий студентам рекомендуется активно работать с имеющимися в библиотеке КБГУ учебниками и учебными пособиями, как бумажными, так и электронными.

Результативность самостоятельной работы студентов во многом определяется наличием активных методов ее контроля, одними из которых являются коллоквиумы, зачеты, рейтинговая система контроля. Использование рейтинговой системы позволяет добиться более ритмичной работы студента в течение семестра, а также активизирует познавательную деятельность студентов путем стимулирования их творческой активности.

### **Методические рекомендации по работе с литературой**

Всю литературу можно разделить на учебники и учебные пособия, оригинальные научные монографические источники, научные публикации в периодической печати. Из них можно выделить литературу основную (рекомендуемую), дополнительную и литературу для углубленного изучения дисциплины.

Изучение дисциплины следует начинать с учебника, поскольку учебник – это книга, в которой изложены основы научных знаний по определенному предмету в соответствии с целями и задачами обучения, установленными программой.

При работе с литературой необходимо учитывать, что имеются различные виды чтения, и каждый из них используется на определенных этапах освоения материала.

Предварительное чтение направлено на выявление в тексте незнакомых терминов и поиск их значения в справочной литературе. В частности, при чтении указанной литературы необходимо подробнейшим образом анализировать понятия.

Сквозное чтение предполагает прочтение материала от начала до конца. Сквозное чтение литературы из приведенного списка дает обучающимся сформировать свод основных понятий из изучаемой области и свободно владеть ими.

Выборочное – наоборот, имеет целью поиск и отбор материала. В рамках данного курса выборочное чтение, как способ освоения содержания курса, должно использоваться при подготовке к практическим занятиям по соответствующим разделам.

Аналитическое чтение – это критический разбор текста с последующим его конспектированием. Освоение указанных понятий будет наиболее эффективным в том случае, если при чтении текстов обучающийся будет задавать к этим текстам вопросы. Часть из этих вопросов сформулирована в ФОС в перечне вопросов для собеседования. Перечень этих вопросов ограничен, поэтому важно не только содержание вопросов, но сам принцип освоения литературы с помощью вопросов к текстам.

Целью изучающего чтения является глубокое и всестороннее понимание учебной информации. Есть несколько приемов изучающего чтения: чтение по алгоритму предполагает разбиение информации на блоки: название; автор; источник; основная идея текста; фактический материал; анализ текста путем сопоставления имеющихся точек зрения по рассматриваемым вопросам; новизна.

Прием постановки вопросов к тексту имеет следующий алгоритм:

- медленно прочитать текст, стараясь понять смысл изложенного;
- выделить ключевые слова в тексте;
- постараться понять основные идеи, подтекст и общий замысел автора.

Прием тезирования заключается в формулировании тезисов в виде положений, утверждений, выводов.

К этому можно добавить и иные приемы: прием реферирования, прием комментирования.

Важной составляющей любого солидного научного издания является список литературы, на которую ссылается автор. При возникновении интереса к какой-то обсуждаемой в тексте проблеме всегда есть возможность обратиться к списку относящейся к ней литературы. В этом случае вся проблема как бы разбивается на составляющие части, каждая из которых может изучаться отдельно от других. При этом важно не терять из вида общий

контекст и не погружаться чрезмерно в детали, потому что таким образом можно не увидеть главного.

Подготовка к экзамену должна проводиться на основе лекционного материала, материала практических занятий с обязательным обращением к основным учебникам по курсу. Это позволит исключить ошибки в понимании материала, облегчит его осмысление, прокомментирует материал многочисленными примерами.

## 8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 8.1. Требования к материально-техническому обеспечению

Для проведения лекционных занятий с компьютерной поддержкой требуется наличие аудитории с проекционным оборудованием, также при изучении дисциплины «Методы обработки больших данных» предполагается использование интерактивной доски.

Во время самостоятельной работы студенты используют компьютерные классы института информатики, электроники и компьютерных технологий, электронные читальные залы КБГУ и домашние компьютеры.

Для проведения лабораторных с компьютерной поддержкой (32 часа) используются компьютерные классы института информатики, электроники и компьютерных технологий.

При проведении занятий лекционного типа используются:

лицензионное программное обеспечение:

- Продукты Microsoft (Desktop EducationALNG LicSaPk OLVS Academic Edition Enterprise) подписка (Open Value Subscription);

- Антивирусное программное обеспечение Kaspersky Endpoint Security Стандартный Russian Edition;

- AltLinux (Альт Образование 8);

свободно распространяемые программы:

- А

с

- WinZip для Windows - программ для сжатия и распаковки файлов;

- А

- F

m

b

Microsoft Reader для Windows – программа для чтения документов в формате PDF

а Для студентов с ограниченными возможностями здоровья созданы специальные условия для получения образования. В целях доступности получения высшего образования по образовательным программам инвалидами и лицами с ограниченными возможностями здоровья университетом обеспечивается (аудитория для самостоятельной работы и коллективного пользования, специальными техническими средствами для обучения инвалидов и лиц с ОВЗ в КБГУ, аудитория № 145 Главный корпус КБГУ):

D 1. Альтернативная версия официального сайта в сети «Интернет» для слабовидящих; License2- Для инвалидов с нарушением зрения (слабовидящих, незрячих): выполнять, анализировать важнейшие информационные ресурсы, предоставлять обучающемуся необходимую помощь, дублирование вслух справочной информации о расписании учебных занятий; наличие средств для усиления остаточного зрения, брайлевской компьютерной техники, видеоувеличителей, программ не визуального доступа к информации, программ-синтезаторов речи и других технических средств приема-передачи учебной информации в доступных формах для студентов с нарушениями зрения;

- задания для выполнения на экзамене зачитываются ассистентом;
- письменные задания выполняются на бумаге, надиктовываются ассистенту обучающимся;

3. Для инвалидов и лиц с ограниченными возможностями здоровья по слуху (слабослышащие, глухие):

- на зачете/экзамене присутствует ассистент, оказывающий студенту необходимую техническую помощь с учетом индивидуальных особенностей (он помогает занять рабочее место, передвигаться, прочитать и оформить задание, в том числе записывая под диктовку);
- зачет/экзамен проводится в письменной форме.

4. Для инвалидов и лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата, созданы материально-технические условия, обеспечивающие возможность беспрепятственного доступа обучающихся в учебные помещения, объекту питания, туалетные и другие помещения университета, а также пребывания в указанных помещениях (наличие расширенных дверных проемов, поручней и других приспособлений).

- письменные задания выполняются на компьютере со специализированным программным обеспечением или надиктовываются ассистенту;
- по желанию студента экзамен проводится в устной форме.

Обучающиеся из числа лиц с ограниченными возможностями здоровья обеспечены электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

## 9. ЛИСТ ПЕРЕУТВЕРЖДЕНИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Рабочая программа:

одобрена на 2024/2024 учебный год. Протокол № \_\_\_\_\_ заседания кафедры от

« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

В рабочую программу внесены следующие изменения:

1. В части раздела «Учебно-методическое и информационное обеспечение дисциплины»

\_\_\_\_\_

\_\_\_\_\_

Разработчик программы \_\_\_\_\_

Зав. кафедрой \_\_\_\_\_

## ПРИЛОЖЕНИЕ

### Распределение баллов текущего и рубежного контроля

№п/п	Вид контроля	Сумма баллов			
		Общая сумма	1-я точка	2-я точка	3-я точка
1	Посещение занятий	до 10 баллов	до 3 б.	до 3б.	до 4б.
2	Текущий контроль:	до 30 баллов	до 10 б.	до 10 б.	до 10 б.
3	Рубежный контроль (тестирование и коллоквиум)	до 30 баллов	до 10 б.	до 10 б.	до 10 б.
4	Итого сумма текущего и рубежного контроля	до 70 баллов	до 23б	до 23 б	до 24 б