

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Кубанский государственный университет»

Факультет компьютерных технологий и прикладной математики
Кафедра вычислительных технологий

УТВЕРЖДАЮ:
Проректор по учебной работе,
качеству образования – первый
проректор _____ Хагуров Т.А.
_____ 05 _____ 2021 г.



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ Б1.О.31 «ОБРАБОТКА БОЛЬШИХ ДАННЫХ»

Направление
подготовки/специальность 02.03.02 **Фундаментальная информатика и
информационные технологии**
(код и наименование направления подготовки/специальности)

Направленность (профиль) /специализация
Математическое и программное обеспечение компьютерных технологий

Программа подготовки академический бакалавриат

Форма обучения очная

Квалификация выпускника бакалавр

Краснодар 2021

Рабочая программа дисциплины «ОБРАБОТКИ БОЛЬШИХ ДАННЫХ» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Программу составил(а):

Приходько Татьяна Александровна, доцент, к. т. н.
Ф.И.О. , должность, ученая степень, ученое звание

подпись

Рабочая программа дисциплины. «ОБРАБОТКИ БОЛЬШИХ ДАННЫХ» утверждена на заседании кафедры Вычислительных технологий протокол №6 от «20» ма1 2021 г. Заведующий кафедрой (разработчика)

Вишняков Ю.М

(фамилия, инициалы

подпись

Утверждена на заседании учебно-методической комиссии факультета Компьютерных Технологий и Прикладной Математики протокол №1 от «20» мая 2021 г

Председатель УМК факультета Коваленко А.В.

фамилия, инициалы

подпись

Рецензенты:

Гаркуша О.В., доцент кафедры информационных технологий ФБГОУ ВО «Кубанский государственный университет», кандидат физико-математических наук.

Схаляхо Ч.А., доцент КВВУ им.С.М.Штеменко, к.ф.-м.н., доцент

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1 Цель освоения дисциплины

Курс «Обработка больших данных» 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.

1.2 Задачи дисциплины

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.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Курс «Обработка больших данных» относится к базовой части блока Б1 дисциплин Дисциплины (модули).

Для изучения дисциплины студент должен владеть знаниями, умениями и навыками по дисциплинам:

Дискретная математика, Алгебраические структуры, Основы программирования, Алгоритмы вычислительной математики, Конструирование алгоритмов и структур данных, Теория алгоритмов и вычислительных процессов, Основы теории вероятностей и статистических методов, Алгоритмы и структуры данных, Математическая логика и теория алгоритмов, Интеллектуальный анализ данных.

Знания, получаемые при изучении дисциплины «Обработка больших данных» используются при изучении профессиональных дисциплин Распределенные задачи и алгоритмы, Программирование в компьютерных сетях, Облачные вычисления, Мультиагентные системы, а также для работ над дипломной и магистерской работой.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующими **профессиональными компетенциями**:

В процессе освоения дисциплины у студента формируются следующие компетенции:
ОПК-2; ОПК-3; ПК-5.

В результате изучения дисциплины у студента формируются:

- представления о феномене больших данных, о научных и технических проблемах и возможностях, связанных с их появлением, о трендах в области технологий хранения и анализа больших данных;
- знания причин возникновения тренда больших данных, процессов анализа больших данных, основных подходов к обработке больших массивов данных, основ языка R;
- умения формулировать алгоритмы в парадигме MapReduce, выбрать подходящий инструмент анализа больших данных, выбрать подходящую технологию хранения больших данных.

Таблица 1. Профессиональные компетенции студента

№ п.п.	Индекс компетенции	Содержание компетенции (или ее части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
1.	ОПК-2	Способен применять компьютерные/супер-компьютерные методы, современное программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности	стандарты обработки и анализа больших данных, и требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных	использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов	способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям
	ОПК-3	Способен к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям.	математические методы анализа данных, методы и прикладные языки для разработки программных решений в области обработки больших данных, математических, информационных и имитационных моделей, для создания информационных ресурсов глобальных сетей	выполнять сбор и анализ данных, в том числе из сети Интернет, производить интерпретацию и оценку полученных результатов	языками системного и прикладного программирования для разработки математических, информационных и имитационных моделей, для обработки информационных ресурсов глобальных сетей и прикладных баз данных.
	ПК-5	Способен применять в	Современные	применить	владеет

		профессиональной деятельности современные языки программирования и методы параллельной обработки данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии.	языки и средства обработки данных (язык R, Python), прикладные библиотеки для анализа данных	современные языки (R и Python) и прикладные библиотеки для анализа данных, сформулировать научную гипотезу и проверить ее достоверность	средствами сбора, обработки и анализа больших данных, средствами оценки эффективности решений
--	--	---	--	---	---

2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 3 зач.ед. (108 часов), их распределение по видам работ представлено в таблице (для студентов ОФО)

Вид учебной работы	Всего часов	Семестры (часы)			
		6			
Контактная работа в том числе:					
Аудиторные занятия (всего):	72,2	72,2			
В том числе:					
Занятия лекционного типа	32	32			
Занятия семинарского типа (семинары, практ. занятия)					
Лабораторные занятия	32	32			
Иная контрольная работа					
Контроль самостоятельной работы	8	8			
Промежуточная аттестация (ИКР)	0,2	0,2			
Самостоятельная работа, в том числе	36	36			
В том числе:					
Курсовая работа					
Проработка учебного (теоретического) материала	10	10			
Выполнение индивидуальных заданий (подготовкосообщений, презентаций)	10	10			
Реферат					
Подготовка к текущему контролю	15,8	15,8			
Контроль:					
Подготовка к экзамену:	-	-			
Общая трудоемкость час	108	108			
в т.ч. контактная работа	72,2	72,2			
зач. ед.	3	3			

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины. Разделы дисциплины, изучаемые в _6_ семестре (очная форма)

№	Наименование разделов	Количество часов		
		Всего	Аудиторная работа	Внеаудиторная работ

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

1	2	3	4	5	6	7
1.	Введение в большие данные. Понятие Data Minig. Прикладные инструменты для работы с Big Data. Технология MapRaduce. Hadoop.	32	8	4	4	16
2.	Технологии анализа данных: Жизненный цикл анализа больших данных, стандарты Когнитивный анализ данных. Визуализация больших данных.	56	16	4	20	16
3.	Технологии хранения больших данных. Распределенные хранилища, NoSql хранилища, классификация и примеры.	19,8	8		8	3,8
	Итого по разделам дисциплины:	107,8	32	8	32	35,8
	ИКР	0,2				
	Итого по дисциплине:	108	32	8	32	35,8

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента, Д-доклад, РГЗ – расчетно-графическое задание.

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа

№ раз-дела	Наименование раздела	Содержание раздела	Форма текущего контроля	Разработ. с участием представителей работодателей
1	2	3	4	5
1	Введение в большие данные. Понятие Data Minig. Прикладные инструменты для работы с Big Data. Технология MapRaduce. Hadoop.	Предпосылки формирования тренда больших данных - Основные вызовы больших данных (4V) - 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).	ЛР	
2	Технологии анализа данных: Жизненный цикл анализа больших данных, стандарты. Когнитивный анализ данных Визуализация больших данных.	Аналитика больших данных. - Процесс аналитики - Стандарты жизненного цикла Big Data: CRISP-DM - Принципы и инструменты аналитики - Задачи и компетенции аналитиков Big Data - Big Data как рынок - Стек технологий	ЛР РГЗ	

		▪ По поддержки принятия решений ▪ Игроки на рынке BD ▪ Крупнейшие проекты BD в России Когнитивный анализ данных ▪ Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки. ▪ Задачи Data Mining и способы их решения. Классификация методов DM. ▪ Области применения DM. ▪ Классы систем DM. ▪ Процесс накопления и анализа данных: Азбука когнитивного анализа. Аналитика больших данных. Математическая статистика Основные понятия статистики и дескриптивный анализ ▪ Шкалы измерений. ▪ Генеральная совокупность и выборка. ▪ Нормальное распределение. Уровень статистической достоверности. ▪ Свойства описательных статистик (Дескриптивный анализ) ▪ Визуальное представление данных ▪ Меры изменчивости Методы DATA MINING ▪ Данные & знания ▪ Типовые задачи Data Mining ▪ Обучаемые и необучаемые задачи ▪ Жизненный цикл проекта DM ▪ Математический аппарат DM ▪ Стандарты DM Методы анализа на графах Случайные графы, безмасштабные графы, социальные сети – сети тесного мира. Закономерности, методы кластеризации на графах.		
		Прикладные инструменты анализа данных. Корреляция. ▪ Готовые комплексные решения: Weka, RapidMiner, Knime, Orange IBM SPSS Modeler (в прошлом Clementine) ▪ Инструменты визуализации: Tableau, Фреймворки на JS, D3 ▪ Корреляция ○ Понятие корреляции ○ Значимость коэффициента корреляции ○ Виды связи между переменными		

3	Технологии хранения больших данных. Распределенные хранилища, NoSql хранилища, классификация и примеры.	Хранилища данных. Регрессия ▪ Хранилища данных ○ OLAP и OLTP системы ○ Характеристики BigData и хранилища данных ○ Почему не реляционные СУБД? ▪ Требования к хранилищам данных ▪ Регрессионный анализ Распределенные базы данных NoSQL. Решение задач Data Mining. Задачи классификации, кластеризации 1. Распределенные базы данных NoSQL ▪ Типы NoSQL ▪ Репликация и шардинг ▪ Пример NoSQL БД Задачи классификации и кластеризации ▪ Desisison Tree ▪ RandomForest ▪ K-means ▪ R и MapReduse Распределенные базы данных NoSQL. Прмеры: HBase, Cassandra, Neo4j, MongoDB. Распределенные файловые системы (РФС). Структура РФС, Требования к РФС, Примеры: HDFS, Google, LustreFS	ЛР Д	
---	---	--	---------	--

2.3.2. Занятия семинарского типа

Занятия семинарского типа – не предусмотрены.

2.3.3. Лабораторные занятия

№ работы	№ раздела дисциплины	Наименование лабораторных работ
1	1	Ознакомление с синтаксисом языка R для анализа данных.(4ч)
2	1	Способы подготовки и отображения данных в R (4ч). Возможности ввода/вывода.
3	2	Решение задач на больших графах (2ч).
4	2	Способы анализа данных в R. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных(2ч).
5	2	Работа с диаграммами и графиками в R (2ч).
6	2	Проверка статистических гипотез (4ч)
7	2	Корреляционный анализ и регрессионный анализ данных (2ч)
8	2	Решение задач Data Mining. Задачи классификации, кластеризации: деревья решений, RandomForest, k-means. (4ч)
9	3	Изучение принципов работы распределенных баз данных
10	3	Развертывание локального кластера Hadoop. Подсчёт слов в тексте, с помощью MapReduce. (4ч)
11	1-3	Круглый стол: Совместное обсуждение результатов РГЗ
12	1-3	Обсуждение итогов курса

2.3.3 Примерная тематика курсовых работ (проектов)

Учебным планом не предусмотрены.

2.3.4 Расчетно-графические задания (индивидуальное задание)

В процессе изучения дисциплины "Обработка больших данных" студентами выполняется одно расчетно-графическое (индивидуальное) задание. Темы заданий для каждого студента различны. Задача РГЗ состоит в проверке умений студентов и проверке эффективности их самостоятельной работы в плане сбора и анализа данных. Темы заданий ежегодно обновляются. Общая тематика соответствует тематике лабораторных работ.

Примеры РГЗ – задания на анализ данных

Подобрать данные для таблицы, приведенной ниже и проанализировать их взаимное влияние, отобразить корреляцию:

- Роста ВВП на прирост населения
- Прироста населения на динамику безработицы
- Прирост людей с высшим образованием на рост промышленного производства
- Прирост людей с высшим образованием на развитие науки
- Прирост людей с высшим образованием на динамику доходов на душу населения
- Динамику безработицы на динамику преступности
- С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.
- 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)	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:

- постановку задачи;
- сформированный набор данных;
- тексты скриптов на языке R;
- результаты тестов на проверку гипотез о корреляции, оценка регрессии, вычисление корреляции в текстовом и графическом виде.
- ясное и подробное пояснение каждого результата, словесную трактовку графиков;
- выводы;
- список использованной литературы.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Раздел 1. Чтение публикаций по истории развития Big Data, Data Mining [3-5]. Изучение языка R и Python [7,8] осн. список, [1-4] - дополнительный.	Приходько Т.А. Методические указания по выполнению лабораторных работ по дисциплине «Обработка больших данных», утвержденные кафедрой вычислительных технологий
2	Раздел 2. Изучение части курса Introduction to Data Science [6], посвященной визуализации. Визуализация стандартных наборов данных при помощи Tableau.	Приходько Т.А. Методические указания по выполнению лабораторных работ по дисциплине «Обработка больших данных», утвержденные кафедрой вычислительных технологий.
3	Раздел 3. Изучение парадигмы Map Reduce. Подсчет кол-ва слов, реализация алгоритма k-means в рамках парадигмы Map Reduce с использованием Hadoop. Чтение публикаций о распределенных хранилищах данных, их особенностях и принципах построения распределенных файловых систем.	Источники основной и дополнительной литературы

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Семестр	Вид занятия (Л, ПР, ЛР)	Используемые интерактивные образовательные технологии	Количество часов
	Л	Компьютерные презентации и обсуждение	34

6	ЛР	Разбор конкретных ситуаций (задач), тренинги по решению задач, компьютерные симуляции (программирование алгоритмов), подготовка и обсуждение докладов.	34
	КРС	Контрольная работа	8
Итого:			72

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1 Фонд оценочных средств для проведения текущего контроля

Фонд оценочных средств дисциплины состоит из средств текущего контроля (вопросы при защите ЛР, контрольной работы) лабораторных работ, средств итоговой аттестации (зачет в 6 семестре).

Оценка успеваемости осуществляется по результатам:

- выполнения лабораторных работ;
- выполнения РГЗ;
- ответа на зачете (для выявления знания и понимания теоретического материала дисциплины).

Текущий контроль включает контрольную работу по итогам первой половины курса.

Пример задания для контрольной работы:

1. Перечислите стандарты жизненного цикла больших данных.
2. Перечислите и охарактеризуйте методы интеллектуального анализа данных.
3. Перечислите классы хранилищ больших данных, назовите различия между ними.

Перечень вопросов, для подготовки к зачету

1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.
2. Дать краткую сравнительную характеристику инструментария ПО для анализа данных.
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.
5. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать 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 оценивания:

"Зачет" - изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;

- наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности.

Практические задания выполнены на 60-100%.

"Не зачет"- баллов (оценка неудовлетворительно) - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы». **Выполнено менее 60% практических заданий.**

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;
- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;
- при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа. Для лиц с нарушениями слуха:
- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

5.1 Основная 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>

5.2. Дополнительная литература:

1. 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>
2. Добронец, Б.С. Численный вероятностный анализ неопределенных данных : монография / Б.С. Добронец, 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&>

5.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Lectures on scientific computing with Python. В свободном доступе: URL: <https://github.com/jrjohansson/scientific-python-lectures>
2. Python. The official Python web site. В свободном доступе: URL: <https://www.python.org/>
3. Программирование и научные вычисления на языке Python В свободном доступе: <http://ru.wikiversity.org/wiki/>
4. Пакет NumPy. Краткое введение: URL: В свободном доступе: <http://pyviy.blogspot.ru/2009/09/numpy.html>
5. Мерков, Александр Борисович. Распознавание образов: введение в методы статистического обучения / А.Б. Мерков; Рос. акад. наук, Ин-т систем. анализа.— Москва: УРСС=URSS, 2010.— 254с.:ил.
(URL:<http://www.recognition.mccme.ru/pub/RecognitionLab.html/slbook.pdf>)
6. Доклад ЦРУ про большие задачи и большие данные: URL: <http://bit.ly/CRUbigdata>
7. Тренажер для освоения основ языка R: <http://tryr.codeschool.com/> от O'Reilly
8. Язык R: из учебной лаборатории — в мир больших данных. Леонид Черняк. osp.ru/os/2012/04/13015768/

6. Методические указания для обучающихся по освоению дисциплины

По курсу предусмотрено проведение лекционных занятий, на которых дается основной систематизированный материал, лабораторных работ, контрольной работы, зачета и экзамена.

Важнейшим этапом курса является самостоятельная работа по дисциплине с использованием указанных литературных источников и методических указаний автора курса.

Виды и формы СР, сроки выполнения, формы контроля приведены выше в данном документе.

Для лучшего освоения дисциплины при защите ЛР студент должен ответить на несколько вопросов из лекционной части курса.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) — дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

7.1 Перечень информационных технологий

- Проверка домашних заданий и консультирование посредством электронной почты.
- Использование электронных презентаций при проведении лекций и практических занятий.

7.2 Перечень необходимого программного обеспечения

1. Python,
2. R, R Studio.
3. Программы для демонстрации и создания презентаций («Microsoft Power Point»).

7.3 Перечень информационных справочных систем:

1. ЭБС Издательства «Лань» <http://e.lanbook.com> ,
2. ЭБС «Университетская библиотека онлайн» www.biblioclub.ru ,
3. ЭБС «Юрайт» <http://www.biblio-online.ru> ,
4. ЭБС «Znaniyum.com» www.znaniyum.com ,
5. ЭБС «Book.ru» <https://www.book.ru>.

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

№	Вид работ	Материально-техническое обеспечение дисциплины
1.	Лекционные занятия	Лекционная аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) и соответствующим программным обеспечением (ПО) PowerPoint. ауд. 129, 131, А305.
2.	Лабораторные занятия	Лаборатория, укомплектованная специализированными техническими средствами обучения – компьютерный класс, с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета. (лаб. 102-106.).
3.	Групповые (индивидуальные) консультации	Аудитория, (кабинет) – компьютерный класс
4.	Текущий контроль, промежуточная аттестация	Аудитория, приспособленная для письменного ответа при промежуточной аттестации.
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.