

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 29.07.2022 17:56:03

Уникальный программный идентификатор:  
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Ростовский государственный экономический  
университет (РИНХ)»

УТВЕРЖДАЮ

Директор Института магистратуры

Иванова Е.А.

« 22 » февраля 2022 г.

**Рабочая программа дисциплины  
Интеллектуальный анализ больших данных**

Направление 01.04.02 Прикладная математика и информатика  
магистерская программа 01.04.02.04 "Искусственный интеллект:  
математические модели и прикладные решения"

Для набора 2022 года

Квалификация  
Магистр

**Составитель программы:**

Рутга Н.А., к.э.н., доц, и.о заведующий кафедрой фундаментальной и прикладной математики

Программа одобрена на заседании кафедры высшей фундаментальной и прикладной математики

«22» февраля 2022 г., протокол № 6

## I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**Цели** освоения дисциплины (модуля): формирование у студентов компетенций в области разработки и использования методов и инструментария обработки и анализа больших массивов данных. Данная цель соотносится с целью образовательной программы в части прикладных решений для систем искусственного интеллекта.

**Задачи:**

- изучение специфики анализа и обработки больших массивов данных;
- овладение основами разработки моделей, методов, алгоритмов для реализации проектов Big Data;
- освоение базовых решений в области построения архитектур больших данных;
- приобретение навыков применения существующих решений в области обработки больших данных.

## II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

2.1. Учебная дисциплина (модуль) Интеллектуальный анализ больших данных относится к части дисциплин (модулей), формируемых участниками образовательных отношений и является обязательной дисциплиной.

2.2. Для изучения данной учебной дисциплины студенту достаточно знаний, полученных в ходе изучения дисциплин: Базы данных, Информационные технологии, Информатика и программирование, Основы нейронных сетей.

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: Обучение с подкреплением, Инженерия знаний, Обработка естественного языка, Компьютерное зрение.

## III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины направлено на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки (специальности): 01.04.02 «Прикладная математика и информатика».

**Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций**

Компетенция	Индикаторы компетенций	Результаты обучения
ОПК-5. Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта	ОПК-5.1. Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта	ОПК-5.1. 3-1. Знает инструментальные среды, программно-технические платформы для решения профессиональных задач ОПК-5.1. У-1. Умеет применять инструментальные среды, программно-технические платформы для решения профессиональных задач

	ОПК-5.2. Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	ОПК-5.2. 3-1. Знает принципы разработки оригинальных программных средств для решения профессиональных задач ОПК-5.2. У-1. Умеет разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта
--	--	--

#### IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов

Форма отчетности: экзамен

#### 4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости Форма промежуточной аттестации
			Контактная работа преподавателя с обучающимися			Самостоятельная работа	
			Лекции	Семинарские (практические занятия)	Лабораторные занятия		
	<b>Раздел 1. Алгоритмы и технологии анализа больших данных</b>	<b>2</b>	<b>12</b>	<b>0</b>	<b>26</b>	<b>21</b>	Тестирование. Выполнение лабораторных работ №1-5
1.	Введение в науку о данных	2	2	0	0	3	
2.	Алгоритмы первичной обработки больших данных	2	2	0	6	3	
3.	Структуры данных в Python и библиотеки для работы с данными в Python	2	2	0	6	3	
4.	Визуализация Big Data	2	2	0	4	6	
5.	Анализ и преобразование больших данных, библиотеки Python	2	2	0	6	3	
6.	Работа с временными рядами в библиотеках Python	2	2	0	4	3	

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Самостоятельная работа	Формы текущего контроля успеваемости Форма промежуточной аттестации
			Контактная работа преподавателя с обучающимися					
			Лекции	Семинарские (практические занятия)	Лабораторные занятия			
	<b>Раздел 2. Хранилища данных и технологии работы с ними</b>	<b>2</b>	<b>16</b>	<b>0</b>	<b>16</b>	<b>24</b>	Выполнение лабораторных работ №6-9	
7.	Системы управления базами данных	2	2	0	0	3		
8.	Использование языка SQL (DML) для создания структур данных	2	2	0	6	3		
9.	Запросы на языке SQL (DML)	2	2	0	6	3		
10.	Объекты баз данных	2	2	0	0	3		
11.	NoSQL хранилища данных	2	2	0	0	3		
12.	Хранилища больших данных класса ключ-значение	2	2	0	4	3		
13.	Документоориентированные хранилища больших данных	2	2	0	0	3		
14.	Графовые хранилища больших данных	2	2	0	0	3		
	<b>Раздел 3. Интеллектуальные платформы анализа больших данных</b>	<b>2</b>	<b>6</b>	<b>0</b>	<b>10</b>	<b>13</b>		Выполнение лабораторной работы №10
15.	Облачные платформы интеллектуального анализа	2	2	0	0	6		

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)			Самостоятельная работа	Формы текущего контроля успеваемости Форма промежуточной аттестации
			Контактная работа преподавателя с обучающимися				
	данных						
16.	Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT	2	2	0	10	3	
17.	Эффективность и отказоустойчивость платформ интеллектуального анализа больших данных	2	2	0	0	4	
<b>Итого</b>		2	34	0	52	58	

#### 4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Затраты времени (час.)		
2	Алгоритмы и структуры данных на Python и SQL	Изучение учебной и научной литературы.	8 недель	28	Реферат.	Рекомендованная учебная литература
2	Машинное обучение и анализ данных	Изучение учебной и научной литературы.	9 недель	30	Реферат.	Рекомендованная учебная литература
Общая трудоемкость самостоятельной работы по дисциплине (час)				58		
Бюджет времени самостоятельной работы, предусмотренный учебным планом для данной дисциплины (час)				58		

### 4.3 Содержание учебного материала

#### Раздел 1. Алгоритмы и технологии анализа больших данных

- 1. Введение в науку о данных.** Анализ данных и математическая статистика: инженерный и научный подходы. Предварительный анализ данных.
- 2. Алгоритмы первичной обработки больших данных.** Классификация данных. Одномерный и многомерный категориальный и количественный анализ данных.
- 3. Структуры данных в Python и библиотеки для работы с данными в Python.** Библиотека Pandas.
- 4. Визуализация Big Data.** Библиотека Matplotlib.
- 5. Анализ и преобразование больших данных, библиотеки Python.** Библиотека NumPY, SciPy, SeaBorn, Scikit Learn.
- 6. Работа с временными рядами в библиотеках Python.** Библиотека Sktime.

#### Раздел 2. Хранилища данных и технологии работы с ними

- 7. Системы управления базами данных.** Основные понятия структуры серверов баз данных.
- 8. Использование языка SQL (DML) для создания структур данных.** Создание логических и физических сущностей баз данных.
- 9. Запросы на языке SQL (DML).** Язык Trasact SQL.
- 10. Объекты баз данных.** Хранимые процедуры, представления, триггеры.
- 11. NoSQL хранилища данных.** Среды проектирования и организации хранения и работы с большими объемами сложно структурированных данных.
- 12. Хранилища больших данных класса ключ-значение.** Хранение линейных структур данных, Redis-хранилище.
- 13. Документоориентированные хранилища больших данных.** Хранилище MongoDB. Интеграция MongoDB с Python с помощью PyMongo.
- 14. Графовые хранилища больших данных.** База данных Neo4j и графовые алгоритмы.

#### Раздел 3. Интеллектуальные платформы анализа больших данных

- 15. Облачные платформы интеллектуального анализа данных.** Платформы MS Azure, Google AI Cloud, Amazon Web-services.
- 16. Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT.**



**17. Эффективность и отказоустойчивость платформ интеллектуального анализа больших данных.** Резервное копирование и восстановление данных. Управление релизами. Безопасность баз данных.

## V. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При проведении лекций и практических занятий используются следующие образовательные технологии:

- мультимедийные лекции
- электронные формы контроля
- самотестирование студентов

Учебный процесс базируется на концепции компетентностного обучения, ориентированного на формирование конкретного перечня профессиональных компетенций, актуализацию получаемых теоретических знаний. Развертывание компетентностной модели обучения предполагает широкое применение инновационных способов организации учебного процесса, в т.ч. применение метода проектного обучения, технологий управляемого самостоятельного обучения в том числе балльно-рейтинговой системы, а также внедрение системы онлайн-поддержки внеаудиторной работы студентов.

Дисциплина может быть реализована частично или полностью с использованием ЭИОС Университета (ЭО и ДОТ). Аудиторные занятия и другие формы контактной работы обучающихся с преподавателем могут проводиться с использованием платформ Microsoft Teams и MOODLE, в том числе, в режиме онлайн-лекций и онлайн- семинаров.

### Наименование тем занятий с использованием активных форм обучения:

№	Тема занятия	Вид занятия	Форма / Методы интерактивного обучения	Кол-во часов
1	Визуализация Big Data	Лабораторная работа	Интерактивные образовательные технологии с применением компьютерной техники	4
2	Запросы на языке SQL (DML)	Лабораторная работа	Интерактивные образовательные технологии с применением компьютерной техники	6
3	Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT	Лабораторная работа	Интерактивные образовательные технологии с применением компьютерной техники	10
Итого часов				20

## VI. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Фонд оценочных средств по дисциплине (модулю) представляет собой комплект оценочных материалов для проведения текущего контроля и промежуточной аттестации обучающихся и оформляется в виде приложения к рабочей программе дисциплины (модуля).

## VII. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 7.1. Основная литература.

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 09.10.2021).

2. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262> (дата обращения: 10.10.2021).

### 7.2. Дополнительная литература.

1. СУБД: язык SQL в примерах и задачах [Электронный ресурс] : учеб. пособие / И.Ф. Астахова [и др.]. — Электрон. дан. — Москва : Физматлит, 2009. — 168 с. — Режим доступа: <https://e.lanbook.com/book/2101>.

2. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2021. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469867> (дата обращения: 10.10.2021).

3. Маркин, А. В. Системы графовых баз данных. Neo4j : учебное пособие для вузов / А. В. Маркин. — Москва : Издательство Юрайт, 2021. — 303 с. — (Высшее образование). — ISBN 978-5-534-13996-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/467452> (дата обращения: 10.10.2021).

4. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / П. Флах. — Электрон. дан. — Москва: ДМК Пресс, 2015. — 400 с. — Режим доступа: <https://e.lanbook.com/book/69955>.

### 7.3. Список авторских методических разработок.

*нет*

### 7.4. Периодические издания (при необходимости)

1. Журнал «Вестник компьютерных и информационных технологий», <http://www.vkit.ru/>

2. Журнал «Информационные технологии», <http://novtex.ru/IT/>

3. Журнал «Известия ЮФУ. Технические науки», <http://izv-tn.tti.sfedu.ru/>

### 7.5. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

1. Научная электронная библиотека. <http://elibrary.ru/>

2. Федеральное хранилище «Единая коллекция цифровых образовательных ресурсов» <http://school-collection.edu.ru/>

3. Университетская библиотека online: [http://biblioclub.ru/index.php?page=main\\_ub\\_red](http://biblioclub.ru/index.php?page=main_ub_red)

MathWorld. <http://mathworld.wolfram.com>

4. Электронно-библиотечная система (ЭБС) ЮРАЙТ [www.biblio-online.ru](http://www.biblio-online.ru).

## VIII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 8.1. Учебно-лабораторное оборудование.

При проведении дисциплины учащиеся должны быть обеспечены:

1. Лекционной аудиторией с мультимедийным презентационным оборудованием для демонстрации презентаций и иллюстративного материала.

2. Аудиторией для лабораторных занятий с аппаратными и программными средствами в соответствии с реализуемой учебной тематикой.

### **8.2. Программное обеспечение.**

8.2.1. Microsoft Windows, Microsoft Office. Договор 232.02.02.03-16/60 от 10.08.2018 г., с 10.08.2018 г. по 10.08.2019 г.; Договор №232.02.02.03-16/46 от 30.08.2019 г., с 31.07.2019 г. по 30.07.2020 г.; Государственный контракт № SC-P/5679-01/07 от 04.12.2007 г., с 21.12.2007 г. (срок использования ПО неограничен)

8.2.2. Python, MySQL, MongoDB, Neo4j – свободно распространяемое ПО.

### **8.3. Технические и электронные средства.**

Комплект учебных презентационных материалов.

## **IX. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Методические указания приведены в учебной литературе, перечисленной в разделе VII.

**УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ (МОДУЛЯ)**  
**Интеллектуальный анализ больших данных**

Трудоемкость: 5 зач.ед.

Форма промежуточной аттестации: экзамен

Курс 1, семестр 2

Код и наименование направления подготовки (специальности): 01.04.02 «Прикладная математика и информатика»

Магистерская программа: «Искусственный интеллект: математические модели и прикладные решения»

№	Виды контрольных мероприятий	Текущий контроль	Рубежный контроль (при наличии)
	<b>Раздел 1. Алгоритмы и технологии анализа больших данных</b>	<b>31</b>	
1.	Выполнение лабораторных работ	25	
2.	Тестирование	6	
	<b>Раздел 2. Хранилища данных и технологии работы с ними.</b>	<b>20</b>	<b>2</b>
1.	Реферат №1		2
2.	Выполнение лабораторных работ	20	
	<b>Раздел 3. Интеллектуальные платформы анализа больших данных</b>	<b>5</b>	<b>2</b>
1.	Реферат №2		2
2.	Выполнение лабораторных работ	5	
	<b>Промежуточная аттестация в форме экзамена</b>	<b>40</b> баллов	<b>Экзамен считается сданным при получении не менее 22 баллов, для допуска к экзамену необходимо набрать не менее 38 баллов по сумме текущего и рубежного контроля. Оценка по дисциплине выставляется по сумме баллов за текущий контроль, рубежный контроль и экзамен: – 85–100 баллов – оценка «отлично»; – 71–84 балла – оценка «хорошо»; – 60–70 баллов – оценка «удовлетворительно»; – менее 60 баллов – оценка «неудовлетворительно»</b>
	<b>ИТОГО</b>	<b>100</b> баллов	

Преподаватель:

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение высшего  
образования «Ростовский государственный экономический университет (РИНХ)»  
Факультет компьютерных технологий и защиты информации  
Кафедра фундаментальной и прикладной математики

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ  
ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ БОЛЬШИХ ДАННЫХ

Код и наименование направления подготовки/специальности:  
01.04.02 «Прикладная математика и информатика»

Уровень образования:  
Магистратура

Магистерская программа:  
«Искусственный интеллект: математические модели и прикладные решения»

Форма обучения:  
Очная

Ростов-на-Дону, 2022

## ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ, ФОРМИРУЕМЫХ ДИСЦИПЛИНОЙ Интеллектуальный анализ больших данных

Код компетенции	Формулировка компетенции
1	2
<b>ОПК</b>	<b>ОБЩЕПРОФЕССИОНАЛЬНЫЕ КОМПЕТЕНЦИИ</b>
ОПК-5.	Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта

### ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ Интеллектуальный анализ больших данных

№ п/п	Контролируемые дисциплины*	разделы	Код контролируемой компетенции	Наименование оценочного средства**
1.		<b>Раздел 1. Алгоритмы и технологии анализа больших данных</b>	ОПК-5	Тестирование, выполнение лабораторных работ
2.		<b>Раздел 2. Хранилища данных и технологии работы с ними</b>		Выполнение лабораторных работ.
3.		<b>Раздел 3. Интеллектуальные платформы анализа больших данных</b>		Выполнение лабораторных работ.

\* Наименование раздела указывается в соответствии с рабочей программой дисциплины.

\*\*Наименование оценочного средства указывается в соответствии с учебной картой дисциплины.

### Тестовые вопросы и ключи

#### по дисциплине «Интеллектуальный анализ больших данных»

1. Data Mining – это процесс обнаружения в сырых данных знаний, необходимых для: расчета целевых показателей; **принятия решений в различных сферах человеческой деятельности**; определения области допустимых значений.

2. Атрибут – это: мера оценки; значение переменной; **свойство, характеризующее объект**

3. В процессе работы Data Mining программы пользователь может получить такие результаты: **только статистически достоверные результаты**; полученные на основе методов математического анализа; расчетные показатели процесса.

4. Данные могут быть получены в результате: **наблюдений; экспериментов; расчетов.**

5. Данные представляют собой: **факты; наблюдения; меры оценки.**

6. Данные – это ... **необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных**; расчетные характеристики процесса; результаты обработки имеющихся параметров.

7. Два основных типа переменных в статистике: **качественные и количественные**; расчетные и априорные; простые и сложные.

8. Для какой шкалы применимы только такие операции как равно и не равно? порядковая шкала; регрессивная шкала; **номинальная шкала**; прогрессивная шкала.

9. Для какой шкалы применимы только такие операции как равно, не равно, больше, меньше? **порядковая шкала**; регрессивная шкала; номинальная шкала; прогрессивная шкала

10. Задачей классификации можно назвать предсказание... **категориальной зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных**; категориальной независимой переменной, основываясь на выборке непрерывных и/или категориальных переменных.

11. Задачей регрессии можно назвать предсказание... **числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных**; текстовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных; числовой независимой переменной, основываясь на выборке непрерывных и/или категориальных переменных

12. Задачи классификации решаются следующими алгоритмами:

- о **нейронные сети**
- о **линейной регрессии**
- о Дейкстры
- о Отжига

13. Задачу классификации нельзя решить с помощью... **алгоритма Apriori**; нейронной сетью; алгоритма линейной регрессии

14. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:

- о **быть практически полезными**
- о **быть объективными**
- о **быть неочевидными**
- о быть точными

15. Изначальная предопределенность классов является характеристикой задачи ... кластеризации; **классификации**; интеграции

16. Инструменты Data Mining работают на основе точных методов; **могут самостоятельно строить гипотезы о взаимосвязях в данных**; позволяют найти глобальный оптимум

17. Интервальная шкала – это шкала **разности между значениями которой могут быть вычислены, однако их отношения не имеют смысла**; суммы значений которые могут быть вычислены, однако их отношения не имеют смысла; произведения значений которые могут быть вычислены, однако их отношения не имеют смысла

18. К какой категории данных относится вес измеряемых объектов: **непрерывным данным**; категориальным данным; основным данным

19. К классу описательных задач Data Mining относятся такие задачи:

- о **кластеризация**
- о поиск ассоциативных правил
- о размещение
- о компоновка

20. К классу прогнозирующих задач Data Mining относятся такие задачи:

- о размещение
- о компоновка
- о регрессия
- о **классификация**

21. Классификация относится к стратегии: **обучения с учителем**; обучение без учителя; обучение учителя

22. Классификация относится к:

- о **управляемому обучению**
- о неуправляемому обучению
- о **контролируемому обучению**
- о обучению по правилам

23. Классификация — это ... отнесение объектов к одному из заранее известных кластеров; **отнесение объектов к одному из заранее известных классов**; отнесение объектов к одной из заранее известных предметных областей

24. Метод деревьев решений применяется для решения задач ... кластеризации; идентификации; интеграции; **классификации**

25. Множество примеров, используемое для конструирования модели, называется... **обучающим множеством**; тестовым множеством; эталонным множеством; предметным множеством

26. Множество примеров, используемое для проверки работы сконструированной модели, называется... **обучающим множеством**; **тестовым множеством**; эталонным множеством; предметным множеством

27. Назовите факторы, обусловившие возникновение и развитие Data Mining:

о совершенствование алгоритмов обработки информации

о **накопление большого количества ретроспективных данных**

о совершенствование технологий хранения и записи данных

о совершенствование аппаратного и программного обеспечения

о появление методов математического анализа

28. Номинальная шкала – это шкала, **содержащая только категории, которые не могут упорядочиваться**; содержащая только категории, которые могут упорядочиваться; содержащая только категории, которые могут интегрироваться

29. Объект описывается как ... **набор атрибутов**; набор переменных; набор правил; набор характеристик

30. Объектом не является: **переменная**; набор переменных; набор правил

31. Основная характеристика задачи бинарной классификации: **зависимая переменная может принимать только два значения**; зависимая переменная может принимать только три значения; зависимая переменная может принимать только четыре значения

32. Оцените правильность утверждения: "Извлечение полезных сведений невозможно без хорошего понимания сути данных" **утверждение верно**; утверждение неверно

33. Оцените правильность утверждения: "Data Mining может заменить аналитика"

о **технология всего лишь дает аналитику инструмент для облегчения и улучшения его работы**

о технология не может дать ответы на те вопросы, которые не были заданы

34. Подготовка данных в процессе Data Mining является: **существенным этапом работы**; подготовительным этапом работы; заключительным этапом работы

35. Порядковая шкала – это шкала, содержащая **категории, которые могут упорядочиваться**; описание предметной области; описание области; описание области допустимых значений

### **Спецификация теста**

Данные тестовые задания предназначены для использования в качестве средства рубежного контроля учебных достижений магистрантов по дисциплине «Интеллектуальный анализ больших данных».

Материалы тестовых заданий предусматривают необходимый минимум проверки знаний по дисциплине, а также степени овладения студентами знаниями в области разработки программно-аппаратных. Тестирование считается успешно пройденным при проценте верных ответов данных обучающимся на более 60% заданий. Успешно сданное тестирование оценивается в 6 баллов.

## **Лабораторные работы по дисциплине**

**по дисциплине «Интеллектуальный анализ больших данных»**



### **Лабораторная работа №1. Алгоритмы первичной обработки больших данных**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №2. Структуры данных в Python и библиотеки для работы с данными в Python**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №3. Визуализация Big Data. Библиотека Matplotlib**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №4. Анализ и преобразование больших данных, библиотеки Python, SciPy, NumPy, Seaborn, Scikit-Learn**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №5. Работа с временными рядами в библиотеках Python.**

#### **Библиотека Stime**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №6. Использование языка SQL (DML) для создания структур данных**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №7. Запросы на языке SQL (DML)**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №8. Объекты баз данных**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №9. Хранилища больших данных класса ключ-значение**

Выполнить задание по варианту, предоставляемому преподавателем.

Предоставить алгоритм и текст программы.

Предоставить отчет о выполненной работе.

### **Лабораторная работа №10. Разработка методов, алгоритмов и программных средств интеллектуального анализа больших данных: Hadoop, Spark и IoT**

Выполнить задание по варианту, предоставляемому преподавателем.  
Предоставить алгоритм и текст программы.  
Предоставить отчет о выполненной работе.

Выполнение одной лабораторной работы оценивается в 5 баллов.

## Темы рефератов

по дисциплине «Интеллектуальный анализ больших данных»

### Часть 1

1. История развития, назначение и роль баз данных.
2. Файловые системы и базы данных.
3. Структуры данных и базы данных.
4. Способы хранения информации в базах данных.
5. Способы повышения эффективности обработки данных за счет их организации.
6. Общая характеристика, назначение, возможности, состав и архитектура СУБД.
7. Классификация СУБД.
8. Информационное, лингвистическое, математическое, аппаратное, организационное, правовое обеспечения СУБД.
9. Типология баз данных. Документальные базы данных. Фактографические базы данных.
10. Типология баз данных. Гипертекстовые и мультимедийные базы данных.
11. Типология баз данных. Объектно-ориентированные базы данных.
12. Типология баз данных. Распределенные базы данных. Коммерческие базы данных.
13. Недостатки реляционных СУБД.
14. Объектные расширения реляционных СУБД.
15. Средства автоматизации проектирования баз данных.
16. Централизация логики приложения на сервере базы данных.
17. Информационные хранилища. OLAP-технология.
18. XML-серверы.
19. Принципы построения БД.
20. Проблема создания и сжатия больших информационных массивов, информационных хранилищ и складов данных.
21. Фрактальные методы в архивации.
22. Управление складами данных.
23. Средства поддержания целостности базы данных
24. Серверы баз данных.
25. Многоплатформенные СУБД. СУБД Oracle.
26. Многоплатформенные СУБД. Informix.
27. Многоплатформенные СУБД. Sybase.
28. Многоплатформенные СУБД. DB2.

### Часть 2

1. Многоплатформенные СУБД. MySQL.
2. СУБД, ориентированные на конкретные платформы. СУБД DBManager в OS/2.
3. СУБД, ориентированные на конкретные платформы. СУБД SQL/400 в AS/400.
4. СУБД, ориентированные на конкретные платформы. СУБД Access в Microsoft Windows.
5. СУБД семейства XBase, Dbase.
6. Базы данных реального времени.
7. Жизненный цикл базы данных.
8. Циклическая база данных.
9. Сжатие без потерь в реляционных СУБД.
10. Защита информации в СУБД.
11. Нормальные формы: НФБК. 3 примера.

12. Нормальные формы: 4НФ. 3 примера.
13. Нормальные формы: 5НФ. Описание. 3 примера.
14. Хранение деревьев в реляционных базах данных.
15. Способы переноса данных с одного типа БД в другую. На примере переноса данных из MySQL в Access.
16. Способы переноса данных с одного типа БД в другую. На примере переноса данных из Access в MySQL.
17. Экспорт/импорт между базами данных различных производителей.
18. Реальные и фантастические разработки БД.
19. Физическое хранение реляционных таблиц.
20. Сериализация транзакций в БД.
21. Анализ качества баз данных.
22. Пути формирования баз данных для директ-маркетинга.
23. Архитектура и функционирование адресных баз данных.
24. Сверхбольшие базы данных.
25. Эксплуатация баз данных. Состав, порядок планирования и проведения регламентных работ.
26. Эксплуатация баз данных. Сервисные средства СУБД.
27. Эксплуатация баз данных. Задачи администратора базы данных.
28. Эксплуатация баз данных. Организация труда обслуживающего персонала.

### **Методические рекомендации по написанию, требования к оформлению**

Рефераты должны отвечать высоким квалификационным требованиям в отношении научности содержания и оформления. Темы рефератов, как правило, посвящены рассмотрению одной проблемы. Объем реферата может быть от 12 до 15 страниц машинописного текста, отпечатанного через 1,5 интервала, а на компьютере через 1 интервал (список литературы и приложения в объем не входят). Текстовая часть работы состоит из введения, основной части и заключения. Во введении студент кратко обосновывает актуальность избранной темы реферата, раскрывает конкретные цели и задачи, которые он собирается решить в ходе своего небольшого исследования. В основной части подробно раскрывается содержание вопроса (вопросов) темы. В заключении кратко должны быть сформулированы полученные результаты исследования и даны выводы. В приложении (приложения) к реферату могут выноситься таблицы, графики, схемы и другие вспомогательные материалы, на которые имеются ссылки в тексте реферата.

### **Критерии оценки:**

- Реферат оценивается в 2 балла, если студент выполнил все предъявляемые требования и рекомендации, в полном объеме раскрыта тема исследования;
- Реферат оценивается в 1 балл при наличии в работе обучающегося серьезных замечаний по содержанию и оформлению работы;
- Реферат оценивается в 0 баллов при ненадлежащем исполнении формы отчетности.

## **Вопросы к экзамену**

по дисциплине «Интеллектуальный анализ больших данных»

1. Какой характер роста данных в технологиях BigData?
2. Развитие каких технологий и устройств привело к появлению термина BigData?
3. Что является основным глобальным источником роста объемов данных?
4. Каковы причины актуальности технологий BigData?
5. Какие этапы включает в себя парадигма BigData?

6. Какие факторы тормозят внедрение технологий BigData?
7. Что представляют собой процедуры ETL?
8. Какой основной недостаток полной автоматизации аналитических процедур?
9. Опишите основные характеристики BigData?
10. Опишите основные методы, используемые для решения задач BigData?
11. Подходы NoSQL.
12. Технология Hadoop.
13. Парадигма MapReduce.
14. Файловая система HDFS.
15. Утилиты Hadoop.
17. Принципы работы с BigData.
18. В чем разница между операторами DELETE и TRUNCATE?
19. Что подразумевается под СУБД? Какие существуют типы СУБД?
20. Что подразумевается под таблицей и полем в SQL?
21. Что такое соединения в SQL?
22. В чем разница между типом данных CHAR и VARCHAR в SQL?
23. Что такое первичный ключ (Primary key)?
24. Что такое ограничения (Constraints)?
25. В чем разница между SQL и MySQL?
26. Что такое уникальный ключ (Unique key)?
27. Что подразумевается под целостностью данных?
27. В чем разница между кластеризованным и некластеризованным индексами в SQL?
29. Напишите SQL-запрос для отображения текущей даты.
30. Что вы подразумеваете под денормализацией?
31. Опишите различные типы индексов.
32. Что такое нормализация и каковы ее преимущества?
33. В чем разница между командами DROP и TRUNCATE?

## Форма экзаменационного билета

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»  
Факультет компьютерных технологий и защиты информации  
Кафедра фундаментальной и прикладной математики

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

По дисциплине «Интеллектуальный анализ больших данных»  
Направление подготовки 01.04.02 «Прикладная математика и информатика»

1. Какой характер роста данных в технологиях BigData?
2. Что подразумевается под СУБД? Какие существуют типы СУБД?

Составитель М.А. Бутакова

И.о. директора МИИ ИМ М.А. Бутакова

### Критерии оценивания

*40 баллов* – ответ полный и правильный; студент хорошо понимает дополнительные вопросы,

*30 баллов* – в ответе допущены две-три ошибки, исправленные после наводящих вопросов преподавателя,

*22 балла* – студент на идейном уровне понимает содержание материала понимает содержание материала, но затрудняется воспроизвести существенные технические детали,

*10 баллов* – студент понимает содержание поставленного в билете вопроса, но слабо ориентируется в содержании основного учебного материала, не может исправить сделанные ошибки при наводящих вопросах преподавателя,

*0 баллов* – ответ отсутствует.

Для успешной сдачи экзамена необходимо набрать 22 балла.