

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Российский государственный гуманитарный университет»  
(ФГБОУ ВО «РГГУ»)

ИНСТИТУТ ИНФОРМАЦИОННЫХ НАУК И ТЕХНОЛОГИИ БЕЗОПАСНОСТИ  
ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ СИСТЕМ И БЕЗОПАСНОСТИ  
Кафедра информационных технологий и систем

## МЕТОДЫ И СРЕДСТВА АНАЛИЗА БОЛЬШИХ ДАННЫХ В КОМПЬЮТЕРНЫХ СЕТЯХ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

09.04.03 Прикладная информатика

---

*Код и наименование направления подготовки/специальности*

Управление данными и знаниями в компьютерных сетях

---

*Наименование направленности (профиля)/ специализации*

Уровень высшего образования: *магистратура*

Форма обучения: *очная, заочная*

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2024

# **МЕТОДЫ И СРЕДСТВА АНАЛИЗА БОЛЬШИХ ДАННЫХ В КОМПЬЮТЕРНЫХ СЕТЯХ**

Рабочая программа дисциплины

Составитель: д-р техн. наук, проф. Н.З Султанов

Ответственный редактор

Зав. кафедрой, к.с.-х.н., доцент Н.Ш. Шукенбаева

**УТВЕРЖДЕНО**

Протокол заседания кафедры  
информационных технологий и систем

№ 8 от 04.03.2024 г.

© Султанов Н.З., 2024

© РГГУ, 2024

## ОГЛАВЛЕНИЕ

1	Пояснительная записка.....	4
1.1	Цель и задачи дисциплины .....	4
1.2	Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций .....	4
1.3	Место дисциплины в структуре основной образовательной программы .....	5
2	Структура дисциплины.....	5
3	Содержание дисциплины .....	6
4	Информационные и образовательные технологии .....	7
5	Оценка планируемых результатов обучения.....	8
5.1	Система оценивания .....	8
5.2	Критерии выставления оценки по дисциплине.....	8
5.3	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине .....	11
6	Учебно-методическое и информационное обеспечение дисциплины.....	13
6.1	Список источников и литературы .....	13
6.2	Перечень ресурсов информационно-телекоммуникационной сети «Интернет» 13	
6.3	Профессиональные базы данных и информационно-справочные системы.....	14
7	Материально-техническое обеспечение дисциплины (модуля).....	14
8	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья.....	15
9	Методические материалы.....	16
9.1	Планы практических занятий .....	16
	Приложение 1. Аннотация рабочей программы дисциплины .....	20

## 1 Пояснительная записка

### 1.1 Цель и задачи дисциплины

Цель дисциплины: изучение методов обработки структурированных и неструктурированных многообразных данных огромных объемов для получения воспринимаемых человеком результатов.

Задачи:

- изучение методов хранения и управления данными формата Big Data;
- изучение методов организации и анализа данных формата Big Data.

### 1.2 Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ПК-1 Способен применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания информационных систем	ПК-1.1. Знает современные методы и инструментальные средства прикладной информатики ПК-1.2. Умеет применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач ПК-1.3. Владеет современными методами и инструментальными средствами прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания ИС	Знать основные принципы и методы хранения, управления, обработки, анализа данных формата Big Data, современные методы и инструментальные средства для работы с большими данными. Уметь строить модели для данных, хранящихся в распределенной файловой системе, применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения задач анализа больших данных. Владеть современными методами прогнозного и инструментальными средствами прогнозного моделирования и анализа данных.
ПК-3 Способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств	ПК-3.1. Знает инновационные инструментальные средства ИТ-сферы ПК-3.2. Умеет проектировать информационные процессы и системы с использованием инновационных инструментальных средств. ПК-3.3. Владеет навыками проектирования информационных	Знать инновационные инструментальные средства ИТ-сферы для работы с большими данными Уметь проектировать информационные процессы и системы с использованием инновационных инструментальных средств. Владеть навыками проектирования информационных процессов и систем с использованием инновационных инстру-

	процессов и систем с использованием инновационных инструментальных средств.	ментальных средств.
ПК-5 Способен использовать современные методы оценки качества, надежности и информационной безопасности информационных систем в процессе их проектирования и эксплуатации	ПК-5.1. Знает современные методы оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации. ПК-5.2. Умеет применять современные методы оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации. ПК-5.3. Владеет навыками применения современных методов оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации прикладных ИС	Знать современные методы оценки качества, надежности и информационной безопасности ИС в процессе работы с большими данными. Уметь применять современные методы оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации систем для работы с большими данными. Владеть навыками применения современных методов оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации ИС класса Big Data.

### 1.3 Место дисциплины в структуре основной образовательной программы

Дисциплина «Методы и средства анализа больших данных в компьютерных сетях» является дисциплиной части, формируемой участниками образовательных отношений блока Б1 учебного плана по направлению подготовки «Прикладная информатика». Имеет индекс Б1.В.05 в структуре учебного плана.

Пререквизиты дисциплины: для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения дисциплин: «Технологии управления знаниями в организации», «Хранилища данных».

Постреквизиты дисциплины: в результате освоения дисциплины формируются знания, умения и владения необходимые для изучения следующих дисциплин: «Облачные технологии», «Распределенные вычисления».

## 2 Структура дисциплины

Общая трудоёмкость дисциплины составляет 4 з.е., 144 академических часа.

### Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
3	Лекции	16
3	Практические работы	24
Всего:		40

Вид итогового контроля – экзамен.

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 86 академических часов, контроль – 18 ч.

Самостоятельная работа включает:

- выполнение индивидуального творческого задания (ИТЗ);
- самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, профессиональных баз данных и информационных справочных систем; подготовка к практическим занятиям; подготовка к текущему контролю и промежуточной аттестации).

### **Структура дисциплины для заочной формы обучения**

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Курс	Тип учебных занятий	Количество часов
2 (зимн. сессия)	Лекции	4
2 (зимн. сессия)	Практические работы	4
2 (летн. сессия)	Лекции	4
2 (летн. сессия)	Практические работы	4
Всего:		16

Вид итогового контроля – экзамен.

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 119 академических часов, контроль – 9 ч.

Самостоятельная работа включает:

- выполнение контрольных работ (КР) в форме индивидуального творческого задания (ИТЗ);
- самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, профессиональных баз данных и информационных справочных систем; подготовка к практическим занятиям; подготовка к текущему контролю и промежуточной аттестации).

### **3 Содержание дисциплины**

№	Наименование раздела дисциплины	Содержание тем
1	Введение в большие данные	<p>1.1 Предметное поле дисциплины. Основная цель изучения дисциплины в структуре ООП ВО. Основные понятия, знания, умения и навыки, получаемые в ходе изучения дисциплины. Структура дисциплины и тематический план. Текущий контроль и промежуточные аттестации. Учебно-методическое обеспечение дисциплины (модуля): основная и дополнительная литература. Периодические издания и интернет-ресурсы.</p> <p>1.2 Предпосылки формирования тренда больших данных. Основные вызовы больших данных (4V).</p>

		<p>Определение термина "большие данные".</p> <p>1.3 Базовое представление о Map Reduce и Hadoop.</p> <p>1.4 Представление о работе аналитика. Процесс аналитики. Принципы аналитики.</p>
2	Введение в Data Mining	<p>2.1 Введение в когнитивный анализ данных. Классификация задач. Функция конкурентного сходства.</p> <p>2.2 Разработка алгоритмов на базе FRIS-функции. Существующие подходы к решению задачи распознавания.</p> <p>2.3 Информативность и выбор признаков. Обнаружение ошибок и заполнение пробелов</p>
3	Основы языка R	<p>3.1 Общие сведения о языке R. Структура языка. Функции. Объекты. Поведение объектов. Выражения. Основные функции. Специальные значения. Приведение типов. Константы (векторы: числовые, буквенные; символы). Операторы (приоритет операций, присвоение). Выражения.</p> <p>3.2 Управляющие структуры (условный оператор; цикл). Структуры данных (индексы: вектор чисел, вектор логических значений, имена). Примитивные типы. Векторы, списки, матрицы, массивы. Таблицы "объект-свойство"</p>
4	Инструменты Data Mining	<p>4.1 Обзор решений. Возможности. Достоинства и недостатки. Области применимости. Weka.</p> <p>4.2 Визуализация. R как инструмент Data Mining. Хранение и доступ к данным по средствам Data Frame. Импорт и экспорт.</p> <p>4.3 Классификация. Регрессия. Кластеризация. R и Hadoop. Основные библиотеки для Data Mining. Возможности библиотеки Pandas.</p>
5	Технологии хранения больших данных	<p>5.1 Зачем нужны новые хранилища. Свойства больших данных и ограничения RDBMS. Определение BigData (3V). Определение BigData (5V+). Скорость, Масштабирование, разнообразие в RDBMS. Структурированность данных.</p> <p>5.2 ACID требования, CAP-теорема, BASE архитектура. Что такое NoSQL. Типы NoSQL. Базы «ключ-значение». Колоночные базы. Документо-ориентированные базы. Графовые базы Интерфейсы NoSQL баз.</p> <p>5.3 Технология распределенных вычислений MapReduce. Упрощенная схема MapReduce. Распределение задания, операции map и reduce.</p>

#### 4 Информационные и образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

## 5 Оценка планируемых результатов обучения

### 5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- защита практических работ	15 баллов	60 баллов
Промежуточная аттестация (экзамен)		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

### 5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>



Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

При оценивании защиты практической работы учитывается:

- полнота выполненной работы (задание выполнено не полностью и/или допущены две и более ошибки или три и более неточности) – 1-7 балла;
- обоснованность содержания и выводов работы (задание выполнено полностью, но обоснование содержания и выводов недостаточны, но рассуждения верны) – 8-12 баллов;
- работа выполнена полностью, в рассуждениях и обосновании нет пробелов или ошибок, возможна одна неточность - 13-15 баллов.

Промежуточная аттестация (экзамен)

При проведении промежуточной аттестации студент должен ответить на 2 вопроса теоретического характера.

При оценивании ответа на вопрос теоретического характера учитывается:

- теоретическое содержание не освоено, знание материала носит фрагментарный характер, наличие грубых ошибок в ответе (1-10 баллов);
- теоретическое содержание освоено частично, допущено не более двух-трех недочетов (11-20 баллов);
- теоретическое содержание освоено почти полностью, допущено не более одного-двух недочетов, но обучающийся смог бы их исправить самостоятельно (21-30 баллов);
- теоретическое содержание освоено полностью, ответ построен по собственному плану (31-40 баллов).

Темы контрольных работ в форме ИТЗ выбираются с учетом предметной области, которая интересует обучающегося, заявок предприятий, в том числе имеющих договора с кафедрой, направления будущей выпускной квалификационной работы. Исходные данные формируются в процессе разработки задания на ИТЗ.

### Оценивание выполнения индивидуального творческого задания

Баллы/ Шкала ECTS	Показатели	Критерии
40...32/ А,В	1. Полнота выполнения индивидуального творческого задания; 2. Своевременность выполнения задания; 3. Последовательность и рациональность выполнения задания; 4. Самостоятельность решения.	Задачи исследования решены самостоятельно, творчески проработан материал. Чётко и последовательно построен доклад и заключение, изложены положения о решении поставленных задач. При этом составлен правильный алгоритм решения; в логических рассуждениях и решении нет ошибок, получено верное заключение, задание выполнено рациональным способом.
33...20/ С		Задачи исследования решены с помощью преподавателя. Не совсем чётко и последовательно построен доклад и заключение, изложены положения о решении поставленных задач. При этом составлен правильный алгоритм решения задания, в логическом рассуждении и решении нет существенных ошибок; есть объяснение решения, но задание выполнено нерациональным способом, получено верное заключение.

Баллы/ Шкала ECTS	Показатели	Критерии
19...10/ D,E		Задачи исследования решены с подсказками преподавателя. Не чётко и последовательно построен доклад и заключение, изложены положения о решении поставленных задач. При этом задание понято правильно, в логическом рассуждении нет существенных ошибок, но допущены существенные ошибки в выборе алгоритма решения; задание выполнено не полностью или в общем виде.
9...0/ F,FX		Задачи исследования не решены.

Затем баллы конвертируются в количество баллов в семестре согласно таблице.

### 5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

#### Вопросы к текущей аттестации

1. Основные вызовы больших данных (4V). (ПК-1, ПК-3)
2. Определение термина "большие данные". (ПК-1, ПК-3)
3. Принципы аналитики. (ПК-1, ПК-3, ПК-5)
4. Процесс аналитики. (ПК-1, ПК-3, ПК-5)
5. Классификация задач. (ПК-1, ПК-3)
6. Функция конкурентного сходства. (ПК-1, ПК-3, ПК-5)
7. Разработка алгоритмов на базе FRiS-функции. (ПК-1, ПК-3)
8. Информативность и выбор признаков. (ПК-1, ПК-3)
9. Обнаружение ошибок и заполнение пробелов. (ПК-1, ПК-3, ПК-5)
10. Структура языка R. (ПК-1, ПК-3)
11. Функции. (ПК-1, ПК-3)
12. Объекты. (ПК-1, ПК-3)
13. Поведение объектов. (ПК-1, ПК-3)
14. Выражения. (ПК-1, ПК-3)
15. Основные функции. (ПК-1, ПК-3)
16. Специальные значения. (ПК-1, ПК-3)
17. Приведение типов. (ПК-1, ПК-3)
18. Константы (векторы: числовые, буквенные; символы). (ПК-1, ПК-3)
19. Операторы (приоритет операций, присвоение). (ПК-1, ПК-3)
20. Управляющие структуры (условный оператор; цикл). (ПК-1, ПК-3)
21. Структуры данных (индексы: вектор чисел, вектор логических значений, имена). (ПК-1, ПК-3)
22. Примитивные типы. (ПК-1, ПК-3)
23. Векторы, списки, матрицы, массивы. (ПК-1, ПК-3)
24. Таблицы "объект-свойство". (ПК-1, ПК-3)
25. Возможности библиотеки Pandas. (ПК-1, ПК-3)
26. Зачем нужны новые хранилища. (ПК-1, ПК-3, ПК-5)
27. Определение BigData (3V). (ПК-1, ПК-3, ПК-5)
28. Определение BigData (5V+). (ПК-1, ПК-3, ПК-5)

29. Скорость в RDBMS (ПК-1, ПК-3)
30. Масштабирование в RDBMS (ПК-1, ПК-3)
31. Разнообразие в RDBMS. (ПК-1, ПК-3)
32. Структурированность данных. (ПК-1, ПК-3)
33. Что такое HBase (ПК-1, ПК-3)
34. Модель данных HBase. Таблица HBase (ПК-1, ПК-3)
35. Свойства документо-ориентированных баз (ПК-1, ПК-3)
36. MongoDB (ПК-1, ПК-3)
37. Репликации и шардинг в MongoDB (ПК-1, ПК-3)
38. Свойства графовых баз данных (ПК-1, ПК-3)
39. Примеры графов (ПК-1, ПК-3)
40. Позиционирование. (ПК-1, ПК-3)
41. Neo4j — что это? (ПК-1, ПК-3)
42. Схема данных Neo4j (ПК-1, ПК-3)
43. Интерфейсы Neo4j (ПК-1, ПК-3)
44. Gephi – open-source инструмент анализа и визуализации сетей. (ПК-1, ПК-3, ПК-5)
45. Запросы в Neo4j (ПК-1, ПК-3)
46. Графовые базы и распределенность. (ПК-1, ПК-3)

#### Вопросы к экзамену

1. Предпосылки формирования тренда больших данных. (ПК-1, ПК-3)
2. Представление о работе аналитика. (ПК-1, ПК-3, ПК-5)
3. Базовое представление о Map Reduce и Hadoop. (ПК-1, ПК-3)
4. Когнитивный анализ данных. (ПК-1, ПК-3, ПК-5)
5. Существующие подходы к решению задачи распознавания. (ПК-1, ПК-3, ПК-5)
6. Общие сведения о языке R. Основные принципы и структура. (ПК-1, ПК-3)
7. Инструменты Data Mining. Обзор решений. (ПК-1, ПК-3)
8. Возможности инструментов Data Mining. Достоинства и недостатки. (ПК-1, ПК-3, ПК-5)
9. Области применимости. Weka. (ПК-1, ПК-3)
10. Визуализация. (ПК-1, ПК-3)
11. R как инструмент Data Mining. (ПК-1, ПК-3)
12. Хранение и доступ к данным по средствам Data Frame. (ПК-1, ПК-3)
13. Импорт и экспорт. (ПК-1, ПК-3)
14. Классификация. (ПК-1, ПК-3)
15. Регрессия. (ПК-1, ПК-3)
16. Кластеризация. (ПК-1, ПК-3)
17. R и Hadoop. (ПК-1, ПК-3)
18. Основные библиотеки для Data Mining. (ПК-1, ПК-3, ПК-5)
19. Свойства больших данных и ограничения RDBMS. (ПК-1, ПК-3, ПК-5)
20. ACID требования. (ПК-1, ПК-3)
21. CAP-теорема. (ПК-1, ПК-3)
22. BASE архитектура. (ПК-1, ПК-3)
23. Что такое NoSQL. (ПК-1, ПК-3, ПК-5)
24. Типы NoSQL. (ПК-1, ПК-3)
25. Базы «ключ-значение». (ПК-1, ПК-3, ПК-5)
26. Колоночные базы. (ПК-1, ПК-3, ПК-5)
27. Документо-ориентированные базы. (ПК-1, ПК-3, ПК-5)
28. Графовые базы. (ПК-1, ПК-3, ПК-5)
29. Интерфейсы NoSQL баз. (ПК-1, ПК-3)

30. Технология распределенных вычислений MapReduce. (ПК-1, ПК-3, ПК-5)
31. Упрощенная схема MapReduce. (ПК-1, ПК-3, ПК-5)
32. Распределение задания, операции map и reduce. (ПК-1, ПК-3)

## **6 Учебно-методическое и информационное обеспечение дисциплины**

### **6.1 Список источников и литературы**

#### **Основная литература**

1. Протоdjяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протоdjяконов, П. А. Пылов, В. Е. Садовников. - Москва ; Вологда : Инфра-Инженерия, 2022. - 392 с. - ISBN 978-5-9729-1006-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1902689> (дата обращения: 19.03.2024). –
2. Зарова, Е. В. Методы Data mining в обработке и анализе статистических данных (решения в R) : монография / Е.В. Зарова. — Москва : ИНФРА-М, 2021. — 232 с. : ил. - ISBN 978-5-16-016814-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1240276>
3. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2020. — 174 с. — URL: <https://urait.ru/bcode/450262>
4. Блануца, В. И. Социально-экономическое районирование в эпоху больших данных: Монография / Блануца В.И. - Москва: НИЦ ИНФРА-М, 2019. - 194 с. - URL: <https://new.znanium.com/catalog/product/1014727>
5. Гуриков, С. Р. Основы алгоритмизации и программирования на Python : учебное пособие / С.Р. Гуриков. — Москва: ФОРУМ : ИНФРА-М, 2020. — 343 с. — (Высшее образование). - ISBN 978-5-00091-487-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1206074>

#### **Дополнительная литература**

1. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2020. — 121 с. — URL: <https://urait.ru/bcode/453758>.
2. Жуков, Р. А. Язык программирования Python: практикум : учеб. Пособие. - М.: ИНФРА-М, 2019. — 216 с. - URL: <https://znanium.com/catalog/product/1000002>.
3. Золотарюк, А. В. Язык и среда программирования R : учебное пособие. - М.: ИНФРА-М, 2020. — 162 с. - URL: <https://znanium.com/catalog/product/1077985>.

### **6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

1. Электронно-библиотечная система «Знаниум» - Режим доступа: <http://znanium.com>
2. - Информационная система «Единое окно доступа к образовательным ресурсам». - Режим доступа: <http://window.edu.ru>
3. Онлайн-энциклопедия. - Режим доступа: <http://encyclopedia.ru>
4. Электронный справочник «Информио» для высших учебных заведений. - Режим доступа: <http://www.informio.ru>
5. КонсультантПлюс. Правовая поддержка. - Режим доступа: <http://www.consultant.ru/>
6. Национальный открытый университет «ИНТУИТ». - Режим доступа: <https://www.intuit.ru/>
7. Сайт Microsoft - Режим доступа: <https://msdn.microsoft.com/ru-ru/library/>
8. Научная библиотека РГГУ - Режим доступа: <http://liber.rsuh.ru/>
9. «CITFORUM»: Аналитическая информация в сфере ИТ. - Режим доступа: <http://citforum.ru/>

### 6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

### 7 Материально-техническое обеспечение дисциплины (модуля)

Для материально-технического обеспечения дисциплины необходимы:

- для лекций:

- учебная аудитория,
- доска,
- проектор (стационарный или переносной),
- компьютер или ноутбук,
- программное обеспечение (ПО).

Перечень программного обеспечения (ПО)

№п/п	Наименование ПО	Способ распространения
1	Microsoft Office 2010 Pro	лицензионное
2	Windows 10	лицензионное
3	Kaspersky Endpoint Security	лицензионное
4	Zoom	лицензионное

- для практических занятий:

- лаборатория,
- доска,
- проектор (стационарный или переносной),
- компьютер или ноутбук для преподавателя,
- компьютеры для обучающихся,
- выход в Интернет,
- программное обеспечение (ПО).

Перечень программного обеспечения (ПО)

Наименование ПО	Способ распространения
Windows 10	лицензионное
Microsoft Office 2010 Pro	лицензионное
Mozilla Firefox	свободно распространяемое
Kaspersky Endpoint Security	лицензионное
Язык программирования R	свободно распространяемое
Microsoft SQL Server	свободно распространяемое
Zoom	лицензионное

Профессиональные полнотекстовые базы данных:

1. Национальная электронная библиотека (НЭБ) [www.rusneb.ru](http://www.rusneb.ru)
2. ELibrary.ru Научная электронная библиотека [www.elibrary.ru](http://www.elibrary.ru)
3. Электронная библиотека Grebennikon.ru [www.grebennikon.ru](http://www.grebennikon.ru)

4. Cambridge University Press
5. ProQuest Dissertation & Theses Global
6. SAGE Journals
7. Taylor and Francis
8. JSTOR

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

## **8 Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья**

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
  - письменные задания выполняются на компьютере в письменной форме;
  - экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;
  - в форме электронного документа;
  - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - передвижными, регулируемые эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.

## **9 Методические материалы**

### **9.1 Планы практических занятий**

#### **Практическая работа № 1 Построение и работа с моделями Data Mining**

Цель работы: знакомство с методами решения задач Data Mining.

Задание

1. Изучить различные алгоритмы для создания моделей Data Mining, позволяющих производить различные виды анализа данных, на основе базы данных AdventureWorksDW.
2. Построение классификатора для индивидуального набора данных при помощи следующих алгоритмов:
  - a. Наивный байесовский классификатор (Naive Bayes);
  - b. Деревья решений (Classification tree).
3. Тестирование эффективности алгоритмов для анализа индивидуального набора данных.

Редактор Data Mining содержит все инструменты и средства отображения для построения и работы с моделями Data Mining. Дополнительная информация по редактору доступна в разделе "Using the Data Mining Tools" в SQL Server Books Online.

Работа со следующими сценариями:

- Целевая отправка писем (Targeted mailing)
- Прогнозирование (Forecasting)



- Рыночная корзина (Market basket)
- Кластеризация последовательностей (Sequence clustering)

В сценарии «Целевая отправка писем» - построение моделей, сравнение их ожидаемых возможностей (при помощи окна Mining Accuracy Chart), а также созданием прогнозов с использованием Prediction Query Builder.

Адресная рассылка

Отдел маркетинга компании Adventure Works заинтересован в увеличении продаж при помощи проведения почтовой кампании, нацеленной на определённых покупателей. Исследуя их характеристики, они хотят обнаружить некоторый шаблон, применимый к потенциальным клиентам, который мог бы впоследствии использоваться для определения наиболее вероятных покупателей.

Кроме того, отдел маркетинга намеревается выявить логические группы среди уже существующих в их базе данных клиентов. Например, группа может содержать покупателей, объединённых по демографическому признаку, обладающих схожим набором покупок.

Adventure Works располагает списком как бывших, так и потенциальных клиентов.

В процессе решения этой задачи, отделу маркетинга потребуется выполнить следующие действия:

- Установить модели, способные выявить наиболее вероятных клиентов из списка потенциальных покупателей
- Кластеризовать своих существующих клиентов

Для выполнения этого сценария потребуется использовать Microsoft Native Bayes, Microsoft Decision Trees, и Microsoft Clustering алгоритмы. Сценарий содержит в себе 5 задач:

- Создание проекта служб Analysis Services
- Создание источника данных
- Создание представления источника данных
- Создание структуры модели Data Mining
- Редактирование Data Mining моделей
- 

Контрольные вопросы

1. Data Mining. Основные понятия и определения. Шкалы. Задачи анализа данных.
2. Классификация. Основные понятия и определения. Правила и деревья классификации. Их соответствие.
3. Классификация. Правила классификации. Алгоритм построения 1-правил. Сверхчувствительность.
4. Классификация. Правила классификации. Алгоритм Naive Bayes.
5. Классификация. Деревья классификации. Алгоритм ID3. Недостатки алгоритма ID3.
6. Классификация. Деревья классификации. Недостатки алгоритма ID3. Алгоритм C4.5.
7. Классификация. Правила и деревья классификации. Алгоритм покрытия.

## Практическая работа № 2 Исследование моделей

Цель работы: рассмотреть возможности SSAS по предоставлению сведений о разработанной модели данных.

После того как модели обработаны, можно просмотреть их, используя закладку Mining Model Viewer в редакторе Data Mining. При помощи выпадающего списка Mining Model в верхней части закладки можно исследовать модели, входящие в структуру.

- Модель Microsoft Decision Trees
- Decision Tree (Дерево решений)
- Сеть зависимостей (Dependency Network)
- Модель Microsoft Clustering
- Кластерная диаграмма (Cluster Diagram)
- Профили кластеров (Cluster Profiles)
- Характеристики кластеров (Cluster Characteristics)
- Cluster Discrimination
- Модель Microsoft Native Bayes
- Сеть зависимостей (Dependency Network)
- Профили атрибута (Attribute Profiles)
- Характеристики атрибута (Attribute Characteristics)
- Attribute Discrimination

#### Контрольные вопросы

1. Для чего предназначены средства по исследованию моделей?
2. Какие сведения они могут предоставить, как это может пригодиться в СППР?
3. Какие из рассмотренных алгоритмов предпочтительны для использования, в каких областях?

### **Практическая работа № 3 Тестирование точности Data Mining-моделей**

Цель работы: выполнить проверку точности моделей.

#### Задание

Теперь модели обработаны и исследованы. Но насколько хорошо они предсказывают ситуацию? Работает ли какая-нибудь из моделей лучше, чем другие?

Используя страницу Mining Accuracy Chart, можно вычислить, как хорошо каждая из моделей предсказывает результат и сравнить их между собой. Этот метод сравнения также иногда называется диаграммой роста (lift chart). На странице The Mining Accuracy Chart использует данные, отделённые от первоисточника, что позволяет сравнивать прогнозы с известными результатами. Затем результаты сортируются и отображаются на графике вместе с идеальной моделью, показывая, насколько хорошо каждая модель делает прогноз. График идеальной модели соответствует теоретической модели, предсказывающей результат со 100% точностью.

#### Создание и использование инструмента диаграммы роста

- Отображение входных столбцов
- Фильтрация входных столбцов
- Диаграмма роста

#### Контрольные вопросы

1. Прокомментируйте полученные результаты
2. Что такое идеальная модель
3. Какая из моделей наиболее близка к идеальной, как вы думаете почему?

### **Практическая работа № 4 Создание прогнозов**

Цель работы Рассмотреть возможности пакета BI по решению задач прогнозирования

#### Задание

Теперь, после того как закончили с моделями анализа, можете перейти к созданию DMX запросов, используя Prediction Query Builder. Prediction Query Builder аналогичен Access Query Builder, где можно использовать drag-and-drop для построения запросов.

- Создание запросов

– Просмотр результатов

Контрольные вопросы

1. Для решения каких практических задач используется прогнозирование?
2. Какие методы прогнозирования вы знаете?

## **АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ**

Дисциплина «Методы и средства анализа больших данных в компьютерных сетях» реализуется на факультете информационных систем и безопасности кафедрой информационных технологий и систем.

Цель дисциплины: изучение методов обработки структурированных и неструктурированных многообразных данных огромных объёмов для получения воспринимаемых человеком результатов.

Задачи:

- изучение методов хранения и управления данными формата Big Data;
- изучение методов организации и анализа данных формата Big Data.

Дисциплина направлена на формирование следующих компетенций:

ПК-1 - способен применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания информационных систем

ПК-3 - способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств

ПК-5 - способен использовать современные методы оценки качества, надежности и информационной безопасности информационных систем в процессе их проектирования и эксплуатации

В результате освоения дисциплины обучающийся должен:

Знать основные принципы и методы хранения, управления, обработки, анализа данных формата Big Data, современные методы и инструментальные средства для работы с большими данными; инновационные инструментальные средства ИТ-сферы для работы с большими данными; современные методы оценки качества, надежности и информационной безопасности ИС в процессе работы с большими данными.

Уметь строить модели для данных, хранящихся в распределенной файловой системе, применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения задач анализа больших данных; проектировать информационные процессы и системы с использованием инновационных инструментальных средств; применять современные методы оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации систем для работы с большими данными.

Владеть современными методами прогнозного и инструментальными средствами прогнозного моделирования и анализа данных; навыками проектирования информационных процессов и систем с использованием инновационных инструментальных средств; применения современных методов оценки качества, надежности и информационной безопасности ИС в процессе проектирования и эксплуатации ИС класса Big Data.

По дисциплине (модулю) предусмотрена промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины (модуля) составляет 5 зачетных единиц.