

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**

### **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

38.03.02 Менеджмент

---

*Код и наименование направления подготовки/специальности*

Менеджмент и цифровая трансформация бизнес-процессов компании

---

*Наименование направленности (профиля)/специализации*

Уровень высшего образования: *бакалавриат*

Форма обучения: *очно-заочная*

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2024

МЕТОДЫ И СРЕДСТВА АНАЛИЗА БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

Составитель:

к.т.н., доцент кафедры информационных технологий и систем Е.Б. Карелина

УТВЕРЖДЕНО:

Протокол заседания кафедры  
информационных технологий и систем  
№ 8 от 04.03.2024

## 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.....	5
4	Образовательные технологии.....	6
5	Оценка планируемых результатов обучения.....	6
5.1	Система оценивания.....	6
5.2	Критерии выставления оценки по дисциплине.....	7
5.3	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине.....	9
6	Учебно-методическое и информационное обеспечение дисциплины.....	10
6.1	Список источников и литературы.....	10
6.2	Перечень ресурсов информационно-телекоммуникационной сети «Интернет» 11	
6.3	Профессиональные базы данных и информационно-справочные системы.....	11
7	Материально-техническое обеспечение дисциплины.....	11
8	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья.....	12
9	Методические материалы.....	14
9.1	Планы семинарских занятий.....	14
	Приложение 1. Аннотация рабочей программы дисциплины.....	17

## 1 Пояснительная записка

### 1.1 Цель и задачи дисциплины

Цель дисциплины: формирование компетенций в области использования информации, обработки и анализа ее для информационно-аналитической поддержки принятия управленческих решений, изучение методов обработки структурированных и неструктурированных многообразных данных огромных объемов для получения воспринимаемых человеком результатов.

Задачи:

- приобретение знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- приобретение практических навыков работы большими данными
- 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
ОПК – 5 Способен использовать при решении профессиональных задач современные информационные технологии и программные средства, включая управление крупными массивами данных и их интеллектуальный анализ.	ОПК – 5.1 Знает современный инструментарий сбора и анализа данных, необходимых для решения поставленных управленческих задач.	Знать: методы извлечения знаний из данных. Уметь: анализировать современные потоки данных, находить, извлекать и структурировать данные. Владеть: современными методами анализа полученных результатов для дальнейшего использования их в решении поставленных управленческих задач.
	ОПК – 5.2 Владеет и эффективно применяет при решении управленческих задач методики цифрового управления и анализа массивов данных.	Знать: основные принципы и методы хранения, управления, обработки, анализа данных формата Big Data. Уметь: строить модели для данных, хранящихся в распределенной файловой системе. Владеть: современными инструментальными средствами прогнозирования и анализа данных в задачах управления бизнес-процессами компании.

### 1.3 Место дисциплины в структуре основной образовательной программы

Дисциплина «Методы и средства анализа больших данных» является дисциплиной части, формируемой участниками образовательных отношений блока Б1 учебного плана.

## 2 Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 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
7	Лекции	12
7	Семинары	12
Всего:		24

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 84 академических часов.

## 3 Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1	Введение в большие данные	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. Процесс аналитики. Принципы аналитики.
2	Введение в Data Mining	Введение в когнитивный анализ данных. Классификация задач. Функция конкурентного сходства. Разработка алгоритмов на базе FRiS-функции. Существующие подходы к решению задачи распознавания. Информативность и выбор признаков. Обнаружение ошибок и заполнение пробелов
3	Основы языка R	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. Управляющие структуры (условный оператор; цикл). Структуры данных (индексы: вектор чисел, вектор логических значений, имена). Примитивные типы. Векторы, списки, матрицы, массивы. Таблицы "объект-свойство"

4	Инструменты Data Mining	Обзор решений. Возможности. Достоинства и недостатки. Области применимости. Weka. Визуализация. R как инструмент Data Mining. Хранение и доступ к данным по средствам Data Frame. Импорт и экспорт. Классификация. Регрессия. Кластеризация. R и Hadoop. Основные библиотеки для Data Mining. Возможности библиотеки Pandas.
5	Технологии хранения больших данных	Зачем нужны новые хранилища. Свойства больших данных и ограничения RDBMS. Определение BigData (3V). Определение BigData (5V+). Скорость, Масштаби-рование, разнообразие в RDBMS. Структурированность данных. ACID требования, CAP-теорема, BASE архитектура. Что такое NoSQL. Типы NoSQL. Базы «ключ-значение». Колоночные базы. Документо-ориентированные базы. Графовые базы Интерфейсы NoSQL баз. Технология распределенных вычислений MapReduce. Упрощенная схема MapReduce. Распределение задания, операции map и reduce.

#### 4 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:		
Решение практических заданий	15 баллов	60 баллов
Промежуточная аттестация – зачет с оценкой		40 баллов
<b>Итого за семестр</b>		<b>100 баллов</b>

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/ А,В	«зачтено (отлично)»	<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 выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«зачтено (хорошо)»	<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 выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«зачтено (удовлетворительно)»	<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 выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		ции. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	не зачтено	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. 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. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

При оценивании практического задания 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 (зачет с оценкой)

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).

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



### Вопросы к текущему контролю успеваемости

1. Основные вызовы больших данных (4V).
2. Определение термина "большие данные".
3. Принципы аналитики.
4. Процесс аналитики.
5. Классификация задач.
6. Функция конкурентного сходства.
7. Разработка алгоритмов на базе FRiS-функции.
8. Информативность и выбор признаков.
9. Обнаружение ошибок и заполнение пробелов.
10. Структура языка R.
11. Функции.
12. Объекты.
13. Поведение объектов.
14. Выражения.
15. Основные функции.
16. Специальные значения.
17. Приведение типов.
18. Константы (векторы: числовые, буквенные; символы).
19. Операторы (приоритет операций, присвоение).
20. Управляющие структуры (условный оператор; цикл).
21. Структуры данных (индексы: вектор чисел, вектор логических значений, имена).
22. Примитивные типы.
23. Векторы, списки, матрицы, массивы.
24. Таблицы "объект-свойство".
25. Возможности библиотеки Pandas.
26. Зачем нужны новые хранилища.
27. Определение BigData (3V).
28. Определение BigData (5V+).
29. Скорость в RDBMS
30. Масштабирование в RDBMS.
31. Разнообразие в RDBMS.
32. Структурированность данных.
33. Что такое HBase.
34. Модель данных HBase. Таблица HBase.
35. Свойства документо-ориентированных баз.
36. MongoDB.
37. Репликации и шардинг в MongoDB.
38. Свойства графовых баз данных.
39. Примеры графов.
40. Позиционирование.
41. Neo4j — что это?
42. Схема данных Neo4j.
43. Интерфейсы Neo4j.
44. Gephi – open-source инструмент анализа и визуализации сетей.
45. Запросы в Neo4j.
46. Графовые базы и распределенность.

**Примеры практических заданий приведены в п. 9.1.**

### Вопросы к зачету с оценкой

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

## 6 Учебно-методическое и информационное обеспечение дисциплины

### 6.1 Список источников и литературы

#### Основная литература

1. Лесковец, Ю. Анализ больших наборов данных : практическое руководство / Д. Дж. Ульман, Ю. Лесковец, А. Раджараман ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2023. - 500 с. - ISBN 978-5-89818-304-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2102592>.
2. Мастицкий, С. Э. Статистический анализ и визуализация данных с помощью R : практическое руководство / С. Э. Мастицкий, В. К. Шитиков. - 2-е изд. - Москва : ДМК Пресс, 2023. - 497 с. - ISBN 978-5-89818-601-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2108480>.
3. Кабаков, Р. R в действии. Анализ и визуализация данных на языке R : практическое руководство / Р. Кабаков ; пер. с англ. П. А. Волковой. - 2-е изд.

- Москва : ДМК Пресс, 2023. - 590 с. - ISBN 978-5-89818-347-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2102634>.

### Дополнительная литература

1. Жукова, Г. С. Математический анализ. Том 1 : учебник / Г. С. Жукова, М. Ф. Рушайло ; под ред. Г. С. Жуковой. — Москва : ИНФРА-М, 2024. — 388 с. — (Высшее образование). — DOI 10.12737/1072169. - ISBN 978-5-16-019247-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2100015>.
2. Мартишин, С. А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем : учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва : ФОРУМ : ИНФРА-М, 2024. — 368 с. — (Высшее образование). - ISBN 978-5-8199-0946-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2096940>.
3. Варфоломеева, Т. Н. Структуры данных и основные алгоритмы их обработки : учебное пособие / Т. Н. Варфоломеева. - 2-е изд., стер. - Москва : ФЛИНТА, 2023. - 159 с. - ISBN 978-5-9765-3691-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2091302>.

## 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 IT. - 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

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	лицензионное
2	Windows	лицензионное
3	Kaspersky Endpoint Security	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	лицензионное
Microsoft Office	лицензионное
Mozilla Firefox	свободно распространяемое
Kaspersky Endpoint Security	лицензионное
Язык программирования R	свободно распространяемое
Microsoft SQL Server	свободно распространяемое

## 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 семинарских занятий**

## Практическое задание № 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 Naïve 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.

#### Практическое задание № 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?

#### Практическое задание № 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?

#### **Практическое задание № 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 реализуется на факультете информационных систем и безопасности кафедрой информационных технологий и систем

Цель дисциплины: формирование компетенций в области использования информации, обработки и анализа ее для информационно-аналитической поддержки принятия управленческих решений, изучение методов обработки структурированных и неструктурированных многообразных данных огромных объемов для получения воспринимаемых человеком результатов.

Задачи:

- приобретение знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- приобретение практических навыков работы большими данными
- изучение методов хранения и управления данными формата Big Data;
- изучение методов организации и анализа данных формата Big Data.

В результате освоения дисциплины обучающийся должен:

*Знать:* методы извлечения знаний из данных, основные принципы и методы хранения, управления, обработки, анализа данных формата Big Data.

*Уметь:* анализировать современные потоки данных, находить, извлекать и структурировать данные, строить модели для данных, хранящихся в распределенной файловой системе.

*Владеть:* современными методами анализа полученных результатов для дальнейшего использования их в решении поставленных управленческих задач, современными инструментальными средствами прогнозного моделирования и анализа данных в задачах управления бизнес-процессами компании.