

МИНОБРНАУКИ РОССИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «ВГУ»)

**УТВЕРЖДАЮ**  
Заведующий кафедрой  
математического моделирования



М.Ш. Бурлуцкая

16.04.2024г.

## **РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

### **Б1.О.05.04 Методы анализа данных**

**1. Код и наименование специальности:**

10.05.04 Информационно-аналитические системы безопасности

**2. Специализация:** Автоматизация информационно-аналитической деятельности;  
Информационная безопасность финансовых и экономических структур

**3. Квалификация выпускника:** Специалист по защите информации

**4. Форма обучения:** Очная

**5. Кафедра, отвечающая за реализацию дисциплины:**

Кафедра математического моделирования

**6. Составитель программы:** Царев Сергей Львович, к.ф.-м.н.

**7. Рекомендована:** Научно-методическим советом математического факультета, протокол  
№ 0500-03 от 28.03.2024

**8. Учебный год:** 2027/2028

**Семестры:** 7, 8

## 9. Цели и задачи учебной дисциплины

### Цели изучения дисциплины:

- изучить основные понятия метода анализа данных, статистические методы анализа данных, особенности хранения и компьютерной обработки информации;
- приобретение навыков и умений построения автоматизированных аналитических моделей.

### Задачи учебной дисциплины:

- овладеть навыками количественного и качественного анализа информации;
- изучение понятийно-категориального аппарата в области углубленного анализа данных;
- знать основные статистические методы анализа данных;
- ознакомиться с алгоритмами компьютерной математики, связанными с саморазвивающимися алгоритмами анализа данных.

## 10. Место учебной дисциплины в структуре ОПОП:

Дисциплина «Методы анализа данных» относится к обязательной части Блока 1. Дисциплины (модули).

## 11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ОПК-10	Способен разрабатывать и применять математические модели и методы анализа массивов данных и интерпретировать профессиональный смысл получаемых формальных результатов	ОПК-10.1	Демонстрирует знания математических моделей и методов анализа массивов данных	<b>Знать:</b> математические модели и методы анализа массивов данных; <b>Уметь:</b> разрабатывать формализованные модели, методы и алгоритмы решения типовых задач автоматизированной информационно-аналитической поддержки процессов принятия решений; решать задачи индивидуального и группового выбора наилучших вариантов решений в условиях неопределенности имеющейся информации на основе использования различных критериев выбора и принципов согласования; <b>Владеть:</b> навыками применения методов машинного обучения для решения задач распознавания, классификации, прогнозирования.
		ОПК-10.2	Разрабатывает формализованные модели, методы и алгоритмы решения типовых задач автоматизированной информационно-аналитической поддержки процессов принятия решений	
		ОПК-10.3	Решает задачи индивидуального и группового выбора наилучших вариантов решений в условиях неопределенности имеющейся информации на основе использования различных критериев выбора и принципов согласования	

## 12. Объем дисциплины в зачетных единицах/час.— 6/216.

Формы промежуточной аттестации: зачет; экзамен.

### 13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость		
		Всего	По семестрам	
			7 семестр	8 семестр
Контактная работа		114	50	64
в том числе:	лекции	66	34	32
	практические	0	0	0
	лабораторные	48	16	32
	курсовая работа			
	контрольные работы			
Самостоятельная работа		66	22	44
Промежуточная аттестация		36		36
Итого:		<b>216</b>	<b>72</b>	<b>144</b>

#### 13.1. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
<b>1. Лекции</b>			
1.1	Введение в предмет	Основы анализа данных. Методология построения моделей сложных систем. Модель черного ящика. Основные этапы моделирования. Методика анализа данных.	
1.2	Методы интеллектуального анализа данных	Определения OLAP, Data Mining, KDD и взаимосвязи между ними. OLAP. Аналитическая отчетность и многомерное представление данных. Хранилище данных. Измерения и факты. Основные операции над кубом данных. Типы задач, решаемые методами Data Mining: классификация, кластеризация, регрессия, ассоциация, поиск последовательных шаблонов. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.	
1.3	Примеры практических приложений в экономике и бизнесе	Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов Data Mining.	
1.4	Подготовка данных и интерпретация результатов	Этапы подготовки данных. Выдвижение гипотез. Методы сбора и систематизации фактов. Методы проведения экспертиз для выявления наиболее значимых факторов. Понятия парциальной и комплексной обработки. Анализ качества полученных моделей.	
1.5	Практические аспекты	Критерии выбора аналитических платформ и пакетов Data Mining. Основные этапы внедрения систем анализа данных. Категории пользователей аналитических систем; требования, предъявляемые к каждой группе пользователей. Способы снижения рисков проектов Data Mining.	
<b>2. Лабораторные занятия</b>			
2.1	Введение в предмет	Основы анализа данных. Методология построения моделей сложных систем. Модель черного ящика. Основные	

		этапы моделирования. Методика анализа данных.	
2.2	Методы интеллектуального анализа данных	Построение аналитической отчетности. Построение регрессионной прогнозной модели спроса.	
2.3	Примеры практических приложений в экономике и бизнесе	Построение скоринговой модели кредитования (деревья решений). Построение нейросетевой прогнозной модели спроса.	
2.4	Подготовка данных и интерпретация результатов	Построение сценария предобработки данных.	
2.5	Практические аспекты	Критерии выбора аналитических платформ и пакетов Data Mining. Способы снижения рисков проектов Data Mining.	

### 13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Введение в предмет	10	-	6	10	26
2	Методы интеллектуального анализа данных	14	-	10	14	38
3	Примеры практических приложений в экономике и бизнесе	16	-	12	16	44
4	Подготовка данных и интерпретация результатов	14	-	12	14	40
5	Практические аспекты	12	-	8	12	32
	<b>Итого:</b>	<b>66</b>	<b>-</b>	<b>48</b>	<b>66</b>	<b>180</b>

### 14. Методические указания для обучающихся по освоению дисциплины:

Освоение дисциплины предполагает не только обязательное посещение обучающимся аудиторных занятий (лекций и лабораторных занятий) и активную работу на них, но и самостоятельную учебную деятельность, на которую отводится 66 часов.

Самостоятельная учебная деятельность студентов по дисциплине «Методы анализа данных» предполагает выполнение следующих заданий:

1) самостоятельное изучение учебных материалов по разделам дисциплины с использованием основной и дополнительной литературы, информационно-справочных и поисковых систем;

2) подготовку к текущим аттестациям: выполнение лабораторных заданий по поиску необходимых для работы в аудитории материалов в Интернете.

Особое внимание обучающихся направляется на освоение математического фундамента дисциплины и её прикладных аспектов.

Вопросы лекционных и практических занятий обсуждаются на занятиях в виде устного опроса – индивидуального и фронтального. При подготовке к лекционным и практическим занятиям, обучающимся важно помнить, что их задача, отвечая на основные вопросы плана занятия и дополнительные вопросы преподавателя, показать свои знания и кругозор, умение логически построить ответ, владение математическим аппаратом и иные коммуникативные навыки, умение отстаивать свою профессиональную позицию. В ходе устного опроса выявляются детали, которые по каким-то причинам оказались недостаточно осмысленными студентами в ходе учебных занятий. Тем самым опрос выполняет важнейшие обучающую, развивающую и корректирующую функции, позволяет студентам учесть недоработки и избежать их при подготовке к промежуточным аттестациям.

Все выполняемые студентами самостоятельно задания (выполнение контрольных и лабораторных работ) подлежат последующей проверке преподавателем. Результаты текущих аттестаций учитываются преподавателем при проведении промежуточной аттестации.

В случае необходимости перехода на дистанционный режим обучения будет создан электронный курс «Методы анализа данных» на портале «Электронный университет ВГУ»: <https://edu.vsu.ru/>. Там же будут размещены необходимые для усвоения курса материалы.

## 15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Технологии анализа данных: DataMining, VisualMining, TextMining, OLAP / А. А. Барсегян [и др.].— 2-е изд. — СПб : БХВ-Петербург, 2007.— 375 с.
2	Тюрин Ю.Н. Анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров ; под ред. В. Э. Фигурнова .— 3-е изд., перераб. и доп. — М. : ИНФРА-М, 2003 .— 543, [1] с.

б) дополнительная литература:

№ п/п	Источник
3	Гайдышев И. Анализ и обработка данных : Спец. справ. / И. Гайдышев. — СПб. и др. : Питер, 2001. — 751 с.
4	Дюк В. Data Mining : Учеб. курс / В. Дюк, А. Самойленко .— СПб. и др. : Питер, 2001 .— 366 с.
5	Саймон Дж. Программирование в Excel : Наглядный курс создания интерактивных электронных таблиц / Дж. Саймон ; Пер. с англ. и ред. Е.Н.Дериевой .— М. : Диалектика, 2002 .— 335 с.

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)\*:

№ п/п	Ресурс
6	Электронный каталог ЗНБ ВГУ : <a href="http://www.lib.vsu.ru">http://www.lib.vsu.ru</a> .
7	<a href="https://math.vsu.ru/wp/?page_id=937">https://math.vsu.ru/wp/?page_id=937</a> – раздел на сайте математического факультета, на котором размещены методические издания.
8	ЭБС «Университетская библиотека онлайн».
9	Электронный университет ВГУ : <a href="https://edu.vsu.ru/">https://edu.vsu.ru/</a> .

## 16. Перечень учебно-методического обеспечения для самостоятельной работы:

№ п/п	Источник
1	Гайдышев И. Анализ и обработка данных : Спец. справ. / И. Гайдышев. — СПб. и др. : Питер, 2001. — 751 с.
2	Дюк В. DataMining : Учеб. курс / В. Дюк, А. Самойленко .— СПб. и др. : Питер, 2001 .— 366 с.
3	Саймон Дж. Программирование в Excel : Наглядный курс создания интерактивных электронных таблиц / Дж. Саймон ; Пер. с англ. и ред. Е.Н.Дериевой .— М. : Диалектика, 2002 .— 335 с.
4	Технологии анализа данных: DataMining, VisualMining, TextMining, OLAP / А. А. Барсегян [и др.].— 2-е изд. — СПб : БХВ-Петербург, 2007.— 375 с.
5	Тюрин Ю.Н. Анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров ; под ред. В. Э. Фигурнова .— 3-е изд., перераб. и доп. — М. : ИНФРА-М, 2003 .— 543, [1] с.
6	Положение об организации самостоятельной работы обучающихся в Воронежском государственном университете.

## 17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ, электронное обучение (ЭО), смешанное обучение):

Дисциплина может реализовываться с применением дистанционных образовательных технологий, например, на платформе «Электронный университет ВГУ» (<https://edu.vsu.ru/>).

Перечень необходимого программного обеспечения: Win10pro или Linux, Microsoft Office, LibreOffice 6, Calc, Microsoft Visual Studio, Microsoft Visual C++, Foxit Reader, браузер Mozilla Firefox, Opera или Internet.

### 18. Материально-техническое обеспечение дисциплины:

Учебная аудитория для проведения занятий лекционного и семинарского типа, текущего контроля и промежуточной аттестации; специализированная мебель.

Для самостоятельной работы используется класс с компьютерной техникой, оснащенный необходимым программным обеспечением, электронными учебными пособиями и законодательно-правовой и нормативной поисковой системой, имеющий выход в глобальную сеть.

### 19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Введение в предмет	ОПК-10	ОПК-10.1, ОПК-10.2	Лабораторная работа № 1
2	Методы интеллектуального анализа данных	ОПК-10	ОПК-10.1, ОПК-10.2, ОПК-10.3	Лабораторная работа № 2
3	Примеры практических приложений в экономике и бизнесе	ОПК-10	ОПК-10.1, ОПК-10.2, ОПК-10.3	Лабораторная работа № 3
4	Подготовка данных и интерпретация результатов	ОПК-10	ОПК-10.1, ОПК-10.2, ОПК-10.3	Лабораторная работа № 4
5	Практические аспекты	ОПК-10	ОПК-10.1, ОПК-10.2, ОПК-10.3	Лабораторная работа № 5
Промежуточная аттестация Форма контроля –зачет; экзамен				Перечень вопросов к зачету; экзамену

### 20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

#### 20.1. Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств: устных опросов, проверки домашних заданий, лабораторных работ, контрольной работы.

#### Лабораторная работа №1

Ответьте на вопросы:

1. Какие из нижеперечисленных признаков являются количественными:

- а) плотность населения
- б) уровень занятости населения
- в) среднедушевой доход
- г) пол человека
- д) возраст
- е) уровень образования (начальное, среднее, высшее)

2. Какие из нижеперечисленных признаков являются дискретными:

- а) денежные доходы населения
- б) число детей в семье
- в) прибыль предприятия

- г) пол человека
- д) тарифный разряд

3. Типологические группировки применяются для:

- а) характеристики структурных сдвигов
- б) характеристики взаимосвязей между отдельными признаками
- в) разделения совокупности на качественно однородные типы
- г) характеристики структуры совокупности

4. Структурные группировки применяются для:

- а) разделения совокупности на качественно однородные типы
- б) характеристики взаимосвязей между отдельными признаками
- в) характеристики структуры совокупности

5. Часть зависимой переменной в регрессионной модели, которая может быть объяснена значением регрессора:

- а) случайное возмущение;
- б) отклик;
- в) уравнение регрессии;
- г) остаток.

## **Лабораторная работа № 2**

1. Гипотеза является сложной, если:

- а) она состоит из конечного числа простых гипотез;
- б) она состоит из бесконечного числа простых гипотез;
- в) Содержит только одно предположение.

2. Коррелированность возмущений с различными номерами называется:

- а) гомоскедастичностью;
- б) гетероскедастичностью;
- в) автокорреляцией.

3. Критической областью называют:

- а) совокупность значений критерия, при которых нулевую гипотезу принимают;
- б) совокупность значений критерия, при которых нулевую гипотезу отвергают.

4. Причины гетероскедастичности (множественный выбор):

- а) исследование неоднородных объектов;
- б) характер наблюдений;
- в) ошибки спецификации;
- г) ошибки измерений.

5. Под мультиколлинеарностью понимается линейная зависимость (единичный выбор):

- а) зависимой переменной с одним или несколькими регрессорами;
- б) двух, или нескольких регрессоров;
- в) зависимой переменной с возмущением;

г) регрессоров с возмущением.

6. С увеличением объема выборки длина доверительного интервала прогнозируемого значения зависимой переменной (единичный выбор):

- а) увеличивается;
- б) уменьшается;
- в) не меняется.

7. Как переводится DATA MINING?

- а) «добыча» или «раскопка данных».
- б) «значение».
- с) «хранение».
- д) «перечисление данных».

8. Какое требование к переработке информации не верно?

- а) Данные имеют неограниченный объем.
- б) Данные являются разнородными.
- с) Результаты должны быть конкретны и понятны.
- д) Инструменты для обработки сырых данных должны быть сложны в использовании.

9. Какая концепция положена в основу современной технологии Data Mining?

- а) Концепция естествознания.
- б) Концепция управления.
- с) Концепция шаблонов (паттернов).
- д) Концепция становления.

10. Сколько выделяют стандартных типов закономерностей?

- а. 4.
- б. 5.
- с. 6.
- д. 7.

11. Если несколько событий связаны друг с другом, то это...

- а) Ассоциация.
- б) Последовательность.
- с) Классификация.
- д) Кластеризация.

12. Основой для каких систем служит историческая информация, хранящаяся в БД в виде временных рядов?

- а) Классификации.
- б) Последовательности.
- с) Прогнозирования.
- д) Ассоциации.

13. Какую иерархическую структуру создают деревья решений?

- а) «ЕСЛИ... ТО...».
- б) «НИ... НИ...».



- c) «КОГДА... ТО...».
- d) «... НИКОГДА...».

14. С чем связано направление эволюционного программирования?

- a) Постановка вопроса вида «значение параметра A больше x?».
- b) Использование метода «ближайшего соседа».
- c) Подача значений входных параметров, на основе которых нужно принимать какие-то решения, прогнозировать развитие ситуации.
- d) Поиск зависимости целевых переменных от остальных в форме функций какого-то определенного вида.

15. Что называют хромосомами в генетических алгоритмах?

- a) Кодировку исходных логических закономерностей в базе данных.
- b) Направление эволюционного программирования.
- c) Большой класс систем.
- d) Набор закономерностей.

### Лабораторная работа № 3

В заданном файле формата XLS содержатся данные финансовых расчетов с потребителями компании за последние полгода. Каждая строка в приведенной базе данных содержит информацию об одной операции отгрузки товара, а именно, имя потребителя, месяц, категорию отгрузки, сумма отгрузки, сумма поступившей оплаты.

А) создайте сводную таблицу для вычисления количества операций отгрузки по каждому потребителю и по каждой категории за весь срок.

Б) создайте сводную таблицу для вычисления общих сумм поставок по каждому потребителю за каждый месяц. Используя полученные данные, постройте соответствующие временные ряды для каждого потребителя.

В) постройте гистограмму (одну) для поступивших оплат только для двух категорий отгрузки «Оборудование» и «Материалы».

### Лабораторная работа № 4

В файле «Дубликаты и противоречия» представлены некоторые исходные данные, а именно столбцы «Вход1», «Вход2», «Выход». В данных присутствуют дубликаты и противоречия. Необходимо построить сценарий, исключающий дубликаты и противоречия.

### Лабораторная работа №5

Коммерческий директор хочет иметь информацию о последних тенденциях в изменении суммовых объемов продаж по товарным группам. Для этого предлагается вычислить относительное изменение объемов продаж за последние 10 дней по отношению к объемам продаж за предыдущие 10 дней. Если это изменение менее -0,3, то товарной группе присваиваем категорию «Провал», от -0,3 до -0,1 – «Падение», от -0,1 до 0 – «Уменьшение», от 0 до 0,1 – «Увеличение», от 0,1 до 0,3 – «Подъем», свыше 0,3 – «Взлет». Необходимо построить соответствующий сценарий.

### Примерный перечень задач для контрольной работы:

1. В чем состоит цель поиска ассоциативных правил? Дайте определения категориям «транзакция», «поддержка набора», «поддержка правила», «достоверность правила».
2. Алгоритм поиска ассоциативных правил Apriori: этапы работы. В чем состоит отличие алгоритма Apriori от алгоритмов AIS и SETM.

3. Организационные факторы при внедрении Data Mining в деятельность компании.
4. Человеческие факторы при внедрении Data Mining в деятельность компании. Основные роли специалистов в процессах Data Mining.
5. Перечислите стандарты методологии Data Mining. В чем состоят их особенности.

Для оценивания текущего контроля успеваемости используются следующие **показатели:**

- 1) знание основных понятий и методов;
- 2) умение применять полученные знания и навыки для решения задач, проводить анализ полученных решений;
- 3) владение математическим аппаратом анализа данных;
- 4) знание имеющихся ресурсов для решения прикладных математических задач;
- 5) умение использовать стандартные пакеты программного обеспечения для решения типовых математических задач;
- 6) владение навыками хранения, поиска, сбора, систематизации, обработки и использования информации.

#### **Шкала оценок:**

Зачтено: Выполнение заданий соответствует перечисленным показателям, обучающийся дает ответы на дополнительные вопросы, может быть не совсем полные. Демонстрирует умение решать задачи, возможно с некоторыми ошибками.

Не зачтено: Ответы не соответствуют ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или их отсутствие.

Для оценивания результатов каждой лабораторной работы используется шкала:

Зачтено: Задания лабораторных работ выполнены верно. Продемонстрировано умение решать задачи, возможно, с не принципиальными ошибками.

Не зачтено: Ответы не соответствуют ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или их отсутствие.

## **20.2. Промежуточная аттестация**

Промежуточная аттестация по дисциплине осуществляется в форме собеседования по экзаменационным билетам с помощью нижеприведенных оценочных средств (перечень вопросов к зачету и экзамену).

В билет включаются теоретический вопрос и задача.

### **Перечень вопросов к зачету:**

№ п/п	Вопросы
1.	Множества данных. Аномальные значения и их влияние на результаты анализа.
2.	ETL-системы. Основные этапы процесса ETL, решаемые им задачи. Очистка данных. Поток данных с точки зрения ETL.
3.	Понижение размерности исходных данных. Свойства алгоритмов понижения размерности.
4.	Трансформация данных, её отличие от предобработки и очистки. Цели трансформации данных в аналитическом приложении.
5.	Системы OLTP и СППР.
6.	Концепция ХД. Основные различия между ХД и обычными базами данных.
7.	Концепция ВХД. Принципы функционирования ВХД. Преимущества и недостатки ВХД по

	сравнению с многомерными иреляционными хранилищами.
8.	Пропущенные значения в данных, необходимость их восстановления. Способы восстановления пропущенных значений. Различие подходов между восстановлением пропусков в упорядоченных и неупорядоченных данных.
9.	Этап загрузки данных в ETL процессе. Основные причины неполной загрузки данных.
10.	Непосредственная загрузка данных из источников.
11.	Особенности извлечения данных из источников разных типов (файлы MS Excel, TXT, СУБД).
12.	Консолидация данных, ее цели и задачи.
13.	Методы очистки и обогащения данных.
14.	Измерения и факты в многомерной модели данных.
15.	Информация и данные. Субъективность информации и объективность данных.
16.	Обогащение данных. Внутреннее и внешнее обогащение данных.
17.	Дубликаты и противоречия.
18.	Принципы построения ХД. Кросс-платформенные ХД, разные архитектуры ХД.
19.	Агрегирование данных.
20.	Метаданные, их роль в процессе функционирования ХД. Виды метаданных.
21.	Качество данных, методы его оценки.
22.	Фиктивные значения данных.
23.	Критические ошибки в данных.
24.	Виды закономерностей в рядах данных.
25.	Детерминированная и случайная составляющая в рядах данных. Шум в данных.
26.	Таблицы фактов и таблицы данных в РХД.
27.	Схемы «звезда» и «снежинка».
28.	Точки среза.
29.	Обучение окнами.
30.	Спектральная обработка данных.
31.	Частотная фильтрация.
32.	Пространственные фильтры.
33.	Профайлинг.
34.	Временные ряды. Одномерные и многомерные временные ряды.
35.	Механизм преобразования данных скользящим окном.
36.	Глубина погружения и горизонт прогноза.
37.	Незначущие и избыточные признаки.
38.	Энтропия, ее использование для отбора признаков.
39.	Метод главных компонент.

### Перечень вопросов к экзамену:

№ п/п	Вопросы
1.	Методы построения моделей сложных систем.
2.	Модель черного ящика.
3.	Основные этапы моделирования.
4.	Методика анализа данных.

5.	Определения OLAP, Data Mining, KDD и взаимосвязи между ними.
6.	Особенности OLAP.
7.	Аналитическая отчетность и многомерное представление данных.
8.	Хранилище данных.
9.	Измерения и факты.
10.	Основные операции над кубом данных.
11.	Типы задач, решаемые методами Data Mining.
12.	Алгоритмы, получившие наибольшее распространение для каждого типа задач:самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети,ассоциативные правила.
13.	Этапы подготовки данных.
14.	Выдвижение гипотез.
15.	Методы сбора и систематизации фактов.
16.	Методы проведения экспертиз для выявления наиболее значимых факторов.
17.	Понятия парциальной и комплексной обработки.
18.	Анализ качества полученных моделей.
19.	Критерии выбора аналитических платформ и пакетов Data Mining.
20.	Основные этапы внедрения систем анализа данных.
21.	Категории пользователей аналитических систем.
22.	Способы снижения рисков проектов Data Mining

#### Примерный перечень задач:

1) Вы работаете в небольшой туристической фирме и планируете массовую рассылку рекламного буклета. Ваши средства ограничены, поэтому вы хотите послать ее тем, кто готов тратить на путешествия и отдых в большей степени. В заданном файле формата XLS содержатся данные о случайной выборке клиентов размером 925 (пол, возраст, суммы, затраченные на путешествия и отдых в предыдущем году). Используйте данные, чтобы понять, насколько пол и возраст влияют на объем затрат. Сформулируйте обоснованные рекомендации относительно контингента для рассылки рекламной брошюры.

2) Основываясь на данных о продажах из файла «Продажи» и других сопутствующих справочниках сформировать сценарий ежедневного отчета по 5 лидерам товарных групп по суммам продаж за последние 15 дней. Отчет визуализировать с помощью столбчатых диаграмм с информацией о сумме продаж и названии товарных групп.

Для оценивания результатов обучения на зачете и экзамене используются следующие **показатели**:

- 1) знание теоретических основ;
- 2) умение решать задачи;
- 3) умение работать с алгоритмами методов и информационными ресурсами;
- 4) успешное прохождение текущей аттестации.

Для оценивания результатов на зачете используется шкала: «зачтено», «не зачтено».

Соотношение показателей, критериев и шкалы оценивания результатов обучения показаны в следующей таблице:

Критерии оценивания	Шкала оценок
Ответ соответствует перечисленным показателям, обучающийся дает ответы на дополнительные вопросы, может быть не совсем полные. Демонстрирует умение решать задачи, возможно с некоторыми ошибками.	«Зачтено»

Ответ не соответствует ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или их отсутствие.	«Не зачтено»
---	--------------

Для оценивания результатов на экзамене используется **шкала**: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Соотношение показателей, критериев и шкалы оценивания результатов обучения показаны в следующей таблице:

Критерии оценивания	Шкала оценок
Ответ соответствует всем перечисленным выше показателям, обучающийся дает ответы на дополнительные вопросы. Демонстрирует знание учебного материала.	«Отлично»
Ответ соответствует двум или более из перечисленных показателей, обучающийся дает ответы на дополнительные вопросы, может быть не совсем полные. Демонстрирует знание учебного материала, возможно с некоторыми ошибками.	«Хорошо»
Ответ соответствует одному из перечисленных показателей, обучающийся не дает ответы на дополнительные вопросы. Демонстрирует знание учебного материала с некоторыми ошибками.	«Удовлетворительно»
Ответ не соответствует ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или отсутствие их.	«Неудовлетворительно»