

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет Управления  
Кафедра «**Бизнес-информатики и высшей математики**»

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **Анализ данных**

Образовательная программа  
**38.03.05 – Бизнес информатика**

Профиль подготовки  
«**Технологическое предпринимательство**»

Уровень высшего образования  
**Бакалавриат**

Форма обучения  
**очная**

Статус дисциплины: **обязательная**

Махачкала, 2019год

Рабочая программа дисциплины Анализ данных составлена в 2019 году в соответствии с требованиями ФГОС ВО по направлению подготовки 38.03.05 Бизнес-информатика (уровень бакалавриата) утвержденного приказом Министерства образования и науки РФ от 11 августа 2016 № 1002

Разработчик: кафедра Бизнес-информатики и высшей математики, к.ф.-м.н. доцент Ибаев Ж.Г.

Рабочая программа дисциплины одобрена:

на заседании кафедры Бизнес-информатики и высшей математики от «20» марта 2019г. протокол № 6  
зав. кафедрой Ибаев Ж.Г. Омарова Н.О.

на заседании Учебно-методической комиссии факультета управления от «10» апреля 2019г. протокол № 8

председатель Гашимова Л.Г. Гашимова Л.Г.

Рабочая программа дисциплины согласована с учебно-методическим управлением «25» апреля 2019г. Гасангаджиева А.Г. Гасангаджиева А.Г.

Уровень высшего образования  
Бакалавриат  
Форма обучения  
Очная  
Статус дисциплины: обязательная

Михайлова  
2019 г.

## СОДЕРЖАНИЕ

<b>Раздел программы</b>	<b>Стр.</b>
Аннотация рабочей программы дисциплины	
1. Цели освоения дисциплины	
2. Место дисциплины в структуре ОПОП бакалавриата	
3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения)	
4. Объем, структура и содержание дисциплины	
5. Образовательные технологии	
6. Учебно-методическое обеспечение самостоятельной работы студентов	
7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины	
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	
10. Методические указания для обучающихся по освоению дисциплины	
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем	
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	

## Аннотация рабочей программы дисциплины

Дисциплина «Анализ данных» входит в базовую часть образовательной программы бакалавриата по направлению 38.03.05 – бизнес информатика

Дисциплина реализуется на факультете управления кафедрой Бизнес-информатики и высшей математики.

Содержание дисциплины охватывает круг вопросов, связанных с анализом данных, изучением способов проведения анализа данных. Рассматриваются методы, алгоритмы и способы анализа данных.

Дисциплина нацелена на формирование следующих компетенций выпускника: общекультурных – ОК-7, общепрофессиональных ОПК-3, профессиональных – ПК-1, ПК-2, ПК-3.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля: текущий контроль успеваемости в форме опросов, рефератов, дискуссий, тестов, решения задач и промежуточный контроль в форме зачета.

Объем дисциплины три зачетных единиц, в том числе в академических часах по видам учебных занятий 108 ч.

Се- мест р	Учебные занятия						СРС, в том чис- ле экза- за- мен	Форма про- межуточной аттестации (зачет, диф- ференциро- ванный за- чет, экзамен
	в том числе							
	Контактная работа обучающихся с преподава- телем							
	Все го	из них						
Лек- ции		Лабора- торные занятия	Практи- ческие занятия	К СР	кон- сульта- ции			
4	108	18	16	16			58	зачет

### 1. Цели освоения дисциплины

Формирование у студентов теоретических знаний, практических умений и навыков по применению современных методов аналитической обработки электронных массивов данных в различных сферах человеческой деятельности.

*Задачи дисциплины:*

изучение существующих технологий подготовки данных к анализу;

изучение основных методов поиска в данных внутренних закономерностей, взаимосвязей, тенденций;

овладение практическими умениями и навыками реализации технологий аналитической обработки данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями;

формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

## 2. Место дисциплины в структуре ОПОП бакалавриата

2 курс (4-й семестр) по очной форме обучения (полный срок обучения). Изучение предусмотрено в базовой части цикла математических и естественнонаучных дисциплин.

Учебная дисциплина «Анализ данных» базируется на изучении дисциплин «Теория вероятностей и математическая статистика», «Имитационное моделирование», «Хранилища данных».

## 3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения)

Компетенции	Формулировка компетенции из ФГОС ВО	Планируемые результаты обучения
<b>ОК -7</b>	способностью к самоорганизации и самообразованию	Знает проблемы и направления научных изысканий в области аналитической обработки данных. Умеет практически применять методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных для качественной подготовки данных к анализу. Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.
<b>ОПК-3</b>	способностью работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных информационных сетях	Знает тенденции разработки универсальных программных средств и аналитических платформ, предназначенных для аналитической обработки данных, построения прогнозов и аналитических сценариев. Умеет применять технологии анализа электронных массивов данных для решения конкретных практических проблем. Владеет навыками выбора и применения отечественных и зарубежных аналитических платформ, используемых для анализа табулированных массивов

		электронных данных.
<b>ПК-1</b>	проводить анализ архитектуры предприятия	Знает алгоритмы поиска ассоциативных правил и кластерного анализа. Умеет практически применять методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных для качественной подготовки данных к анализу. Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.
<b>ПК-2</b>	проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий	Знает статистические и машинные методы классификации и регрессии. Умеет применять технологии анализа электронных массивов данных для решения конкретных практических проблем. Владеет навыками выбора и применения отечественных и зарубежных аналитических платформ, используемых для анализа табулированных массивов электронных данных.
<b>ПК-3</b>	выбор рациональных информационных систем и информационно - коммуникативных технологий решения для управления бизнесом	Знает методики анализа и прогнозирования временных рядов. Умеет использовать возможности отечественных и зарубежных универсальных программных средств и аналитических платформ для аналитической обработки данных, построения прогнозов и аналитических сценариев. Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.

#### **4. Объем, структура и содержание дисциплины.**

4.1. Объем дисциплины составляет 3 зачетных единиц, 108 академических часов.

4.2. Структура дисциплины.

№ п/п	Разделы и темы дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Самостоятельная работа	Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные	Контроль самост.		
<b>Модуль 1. (Вводный модуль)</b>									
1	Аффинитивный анализ. Поиск последовательных шаблонов.	4		2	2	2		8	
2	Кластерный анализ.	4		2	2	2		8	
	<i>Итого по модулю 1:</i>			4	4	4		16	к/р по модулю №1
<b>Модуль 2. (Базовый модуль)</b>									
1	Классификация и регрессия. Статистические методы.	4		4	4	4		10	
2	Классификация и регрессия. Машинное обучение	4		4	4	4		12	
	<i>Итого по модулю 2:</i>			8	8	8		22	к/р по модулю №2
<b>Модуль 3. (Дополнительный модуль)</b>									
1	Анализ и прогнозирование временных рядов.	4		4	2	2		10	
2	Ансамбли моделей.	4		2	2	2		10	
	<i>Итого по модулю 3:</i>			6	4	4		20	к/р по модулю №1
	<b>ИТОГО:</b>	4		18	16	16		58	зачет

### 4.3. Содержание дисциплины, структурированное по темам (разделам).

#### 4.3.1. Содержание лекционных занятий по дисциплине.

## **Модуль 1: Вводный модуль**

### **Тема 1. Аффинитивный анализ. Поиск последовательны шаблонов**

Введение в аффинитивный анализ (affinity analysis). Алгоритм apriori. Иерархические ассоциативные правила.

### **Тема 2. Кластерный анализ**

Введение в кластеризацию. Классификация методов кластеризации. Алгоритм кластеризации k-means. Сети Кохонена (KCN – Kohonen network). Карты Кохонена (SOM – self organizing map). Проблемы алгоритмов кластеризации.

## **Модуль 2: Базовый модуль**

### **Тема 3. Классификация и регрессия. Статистические методы.**

Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель. Множественная линейная регрессия. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Основы логистической регрессии. Интерпретация модели логистической регрессии. Множественная логистическая регрессия.

### **Тема 4. Классификация и регрессия. Машинное обучение**

Введение в деревья решений. Алгоритмы построения деревьев решений. Алгоритмы ID3 и C4.5. Алгоритм CART. Упрощение деревьев решений. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки.

## **Модуль 3: Дополнительный модуль**

### **Тема 5. Анализ и прогнозирование временных рядов.**

Временной ряд и его компоненты. Модели прогнозирования. Прогнозирование в торговле и логистике

### **Тема 6. Ансамбли моделей.**

Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.

### ***4.3.2. Содержание практических занятий по дисциплине.***

## **Модуль 1. Вводный модуль**

### *Тема 1. Аффинитивный анализ. Поиск последовательны шаблонов*

Вопросы к теме:

1. Значение, цели и задачи анализа деятельности предприятий.
2. Предмет и последовательность проведения анализа деятельности предприятий.



3. Информационное и методическое обеспечение анализа деятельности предприятий.

4. Системный подход в анализе данных.

*Ссылка на учебно-методическую литературу, указанную в п.8(1,2,3,4,7,8,9, 10)*

*Тема 2. Кластерный анализ*

1. Понятие кластеризации данных.
2. Методы кластеризации.
3. Алгоритмы кластеризации k-means.
4. Сети Кохонена (KCN – Kohonen network).
5. Карты Кохонена (SOM – self organizing map).
6. Проблемы алгоритмов кластеризации.

*Ссылка на учебно-методическую литературу, указанную в п.8(1,2,3,4,7,8,9, 10)*

## **Модуль 2. Базовый модуль**

*Тема 3. Классификация и регрессия. Статистические методы.*

1. Понятия классификации и регрессии.
2. Простая линейная регрессия.
3. Оценка соответствия простой линейной регрессии реальным данным.
4. Простая регрессионная модель.
5. Множественная линейная регрессия.
6. Модель множественной линейной регрессии.
7. Регрессия с категориальными входными переменными.
8. Методы отбора переменных в регрессионные модели.
9. Ограничения применимости регрессионных моделей.
10. Основы логистической регрессии.
11. Интерпретация модели логистической регрессии.
12. Множественная логистическая регрессия.

*Ссылка на учебно-методическую литературу, указанную в п.8 (1,2,3,4,7,8,9, 10)*

*Тема 4. Классификация и регрессия. Машинное обучение*

1. Введение в деревья решений.
2. Алгоритмы построения деревьев решений.
3. Алгоритмы ID3 и C4.5.
4. Алгоритм CART.
5. Упрощение деревьев решений.
6. Введение в нейронные сети.
7. Искусственный нейрон.
8. Принципы построения нейронных сетей.
9. Алгоритмы обучения нейронных сетей.
10. Алгоритм обратного распространения ошибки.

*Ссылка на учебно-методическую литературу, указанную в п.8(1,2,3,4,7,8,9)*

### **Модуль 3. Дополнительный модуль**

*Тема 5. Анализ и прогнозирование временных рядов.*

1. Временной ряд и его компоненты.
2. Модели прогнозирования.
3. Прогнозирование в торговле и логистике

*Ссылка на учебно-методическую литературу, указанную в п.8  
(1,2,3,4,7,8,9, 10)*

*Тема 6. Ансамбли моделей.*

1. Понятие ансамбля моделей.
2. Бэггинг.
3. Бустинг.
4. Альтернативные методы построения ансамблей.

*Ссылка на учебно-методическую литературу, указанную в п.8  
(1,2,3,4,7,8,9, 10)*

## **5. Образовательные технологии**

Современные образовательные технологии в преподавании дисциплины «Анализ данных» ориентированы на реализацию инновационных методов обучения как слагаемых учебного процесса. Они учитывают преимущества компетентностного подхода к изучению дисциплины, обеспечивают повышение качества знаний, необходимых для профессиональной деятельности управленцев.

Для проведения лекционных и практических занятий используются различные образовательные технологии с использованием активных и интерактивных форм обучения.

Лекции проводятся с использованием средств визуализации лекционного материала (мультимедийных презентаций) и применением таких методов и технологий, как тематическая дискуссия, проблемная лекция, обзорная лекция, лекция-информация с использованием фактических данных и т.п.

При ведении практических занятий по данной дисциплине используются такие стандартные методы обучения, как тестирование, фронтальный опрос, индивидуальный опрос, выполнение лабораторных работ, метод малых групп и т.п. При проведении практических занятий в интерактивной форме используются следующие методы: анализ конкретных ситуаций, тематическая групповая дискуссия, деловая игра, блиц-опрос, проблемный семинар, научный кружок.

К участию в проведении занятий привлекаются также представители российских компаний, государственных и общественных организаций.

## **6. Учебно-методическое обеспечение самостоятельной работы студентов.**

Для теоретического и практического усвоения дисциплины большое значение имеет самостоятельная работа студентов, она осуществляется студентами индивидуально и под руководством преподавателя.

Самостоятельная работа по дисциплине, предусмотренная учебным планом в объеме 58 часов, направлена на более глубокое усвоение изучаемого курса, формирование навыков исследовательской работы и ориентирование студентов на умение применять теоретические знания на практике.

Основными видами самостоятельной работы студентов в рамках освоения дисциплины «Анализ данных» выступают следующие:

- 1) проработка учебного материала;
- 2) работа с электронными источниками;
- 3) обработка аналитических данных;
- 4) подготовка докладов к участию в тематических дискуссиях;
- 5) работа с вопросами;
- 6) написание рефератов;

*Виды и формы контроля самостоятельной работы студентов в рамках освоения дисциплины «Анализ данных»*

Разделы дисциплины	Виды самостоятельной работы (и ссылки на литературу <sup>1</sup> )	Количество часов	Форма контроля
<u>Раздел 1.</u> Аффинитивный анализ. Поиск последовательных шаблонов. Кластерный анализ.	проработка учебного материала, работа с электронными источниками, решение задач, выполнение лабораторных работ, обработка аналитических данных, подготовка докладов к участию в тематических дискуссиях, работа с тестами и вопросами, написание рефератов. (1,2,3,4,7,8,9, 10)	18	Тестирование, дискуссия, опрос,
<u>Раздел 2.</u> Классификация и регрессия. Статистические методы. Классификация и регрессия. Машинное обучение	проработка учебного материала, работа с электронными источниками, решение задач, выполнение лабораторных работ, обработка аналитических данных, подготовка докладов к участию в тематических дискуссиях, работа с тестами и вопросами, написание рефератов. (1,2,3,4,7,8,9)	22	Тестирование, дискуссия, опрос, проверка домашнего задания, обсуждение докладов, защита рефератов
<u>Раздел 3.</u> Анализ и прогнозирование временных рядов. Ансамбли моделей	проработка учебного материала, работа с электронными источниками, решение задач, лабораторных работ, обработка аналитических данных, подготовка докладов к участию в тематических дискуссиях, работа с тестами и вопросами, написание рефератов.	18	Тестирование, дискуссия, опрос, проверка домашнего задания, обсуждение докладов, защита

<sup>1</sup> Дается ссылка на учебно-методическую литературу, указанную в п. 8.

	(1,2,3,4,7,8,9, 10)		рефератов
Итого		58	

Написание реферата используется в учебном процессе с целью развития у студентов умения и навыков самостоятельного научного поиска: изучения литературы по выбранной теме, анализа различных источников и точек зрения, обобщения материала, выделения главного, формулирования выводов. С помощью рефератов студент глубже изучает разделы и темы дисциплины, учится логически мыслить, оформлять, докладывать, презентовать и защищать результаты самостоятельно проведенного научного исследования.

Процесс подготовки, написания и защиты реферата включает:

- выбор темы;
- подбор специальной литературы и иных источников, их изучение;
- составление плана;
- написание и оформление текста (5-15 машинописных страниц);
- подготовка тезисов доклада (на 7-10 минут);
- устное изложение в виде доклада, в том числе виде презентации.

Выбор темы реферата осуществляется в соответствии с предложенной преподавателем тематикой. В отдельных случаях студент может выбрать для своего реферата тему в соответствии с направлением его НИР.

Материал в реферате располагается в следующей последовательности:

- титульный лист;
- план работы;
- введение;
- текст работы (разбитый на разделы);
- заключение
- список литературы.

Содержание реферата студент докладывает на практическом занятии, заседании научного кружка, научно-практической конференции. На основе обсуждения студенту выставляется соответствующий балл за СРС.

### ***Примерная тематика рефератов по модулям и темам дисциплины***

#### **Модуль 1. Вводный модуль**

##### *Тема 2. Кластерный анализ*

1. Дисперсионный анализ данных.
2. Кластерный анализ данных.
3. Методы классификации в Data mining.
4. Регрессионный анализ данных.
5. Анализ данных с использованием бинарной логистической регрессии.
6. Факторный анализ данных.
7. Метрики, применяемые в Data mining.

#### **Модуль 2. Базовый модуль**

*Тема 3. Классификация и регрессия. Статистические методы*

1. Ковариационный анализ данных.
2. Методы поиска ассоциативных правил.
3. Сиквенциальный анализ данных (поиск последовательных шаблонов).
4. Основные стандарты Data mining.
5. Анализ данных с использованием сети Кохонена.
6. Характеристики инструментальных средств Data mining.
7. Реляционные хранилища данных.

*Тема 4. Классификация и регрессия. Машинное обучение*

1. Многомерные хранилища данных.
2. Гибридные хранилища данных.
3. Виртуальные хранилища данных.
4. Характеристика ETL-процесса.
5. Оценка качества, очистка и предобработка анализируемых данных.
6. Сокращение размерности исходного множества анализируемых данных.
7. Искусственные нейронные сети. Многослойный персептрон.

### **Модуль 3. Дополнительный модуль**

*Тема 5. Анализ и прогнозирование временных рядов*

1. Анализ данных с использованием генетических алгоритмов.
2. Анализ данных с использованием самоорганизующихся карт.
3. Оценка значимости регрессионных моделей с применением t-критерия Стьюдента
4. Оценка значимости регрессионных моделей с применением F-критерия Фишера.
5. Алгоритм построения деревьев решений ID3
6. Алгоритм построения деревьев решений C4.5.
7. Оценка полезности, эффективности и точности моделей, применяемых для анализа данных.

*Тема 6. Ансамбли моделей*

1. Анализ данных с использованием ансамблей моделей.
2. Проблемы обучения и переобучения моделей.
3. Технологии обогащения данных.
4. Повышение эффективности моделей с помощью бэггинга и бустинга.
5. Градиентный алгоритм обучения многослойного персептрона (алгоритм обратного распространения ошибки).
6. Lift и Profit-кривые.
7. ROC-анализ.

**7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.**

### 7.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

Перечень компетенций с указанием этапов их формирования приведен в описании образовательной программы.

Код компетенции из ФГОС ВО	Наименование компетенции из ФГОС ВО	Планируемые результаты обучения	Процедура освоения
ОК -7	способностью к самоорганизации и самообразованию	<p>Знает проблемы и направления научных изысканий в области аналитической обработки данных.</p> <p>Умеет практически применять методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных для качественной подготовки данных к анализу.</p> <p>Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.</p>	Устный опрос, письменный опрос, дискуссия, решение задач
ОПК-3	способностью работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных информационных сетях	<p>Знает тенденции разработки универсальных программных средств и аналитических платформ, предназначенных для аналитической обработки данных, построения прогнозов и аналитических сценариев.</p> <p>Умеет применять технологии анализа электронных массивов данных для решения конкретных практических проблем.</p> <p>Владеет навыками выбора и применения отечественных и зарубежных аналитических платформ, используемых для анализа табулированных массивов электронных данных.</p>	Устный опрос, письменный опрос, дискуссия, решение задач
ПК-1	проводить анализ архитектуры предприятия	<p>Знает основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных;</p> <p>Умеет использовать возможности отечественных и</p>	Устный опрос, письменный опрос, дискуссия, решение задач

		зарубежных универсальных программных средств и аналитических платформ для аналитической обработки данных, построения прогнозов и аналитических сценариев; Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.	
<b>ПК-2</b>	проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий	Знает принципы построения и структурную организацию хранилищ данных; Умеет свободно ориентироваться на современном динамичном рынке аналитических программных продуктов. Владеет навыками выбора и применения отечественных и зарубежных аналитических платформ, используемых для анализа табулированных массивов электронных данных.	Устный опрос, письменный опрос, дискуссия, решение задач
<b>ПК-3</b>	выбор рациональных информационных систем и информационно-коммуникативных технологий решения для управления бизнесом	Знает алгоритмы поиска ассоциативных правил и кластерного анализа. Умеет практически применять методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных для качественной подготовки данных к анализу. Владеет технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области.	Устный опрос, письменный опрос, дискуссия, решение задач

## 7.2. Типовые контрольные задания

### *Примерная тематика докладов*

1. Практическое применение алгоритмов Data mining.
2. Классификация с несколькими независимыми переменными методом Naïve Bayes.
3. Поиск оптимальной функции методом наименьших квадратов.

4. Сиквенциальный анализ.
5. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации.
6. Кластеризация данных при помощи нечетких отношений.
7. Характеристика классов задач, решаемых методами Data Mining.
8. Стандарты Data mining. Характеристика стандартов CWM и PMLL.
9. Библиотеки доступа к алгоритмам Data mining. Характеристика библиотеки Xelopes.
10. Характеристика программных инструментов для выполнения интеллектуального анализа данных.
11. Общая характеристика и классификация методов кластерного анализа данных.
12. Анализ данных с использованием методов классификации и регрессии.
13. Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.
14. Сферы применения деревьев решений.
15. Цели, задачи и принципы работы нейронных сетей.
16. Алгоритмы обучения нейронных сетей.
17. Цели, задачи, принципы и модели прогнозирования..

***Примерные вопросы для подготовки к промежуточной аттестации по итогам освоения дисциплины (зачет, 4 семестр)***

1. Модели и их свойства. Аналитический и информационный подходы к моделированию.
2. Формы представления, типы и виды анализируемых данных.
3. Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Ошибки обучения. Эффект переобучения.
4. Общая схема анализа данных. Требования к алгоритмам анализа данных.
5. Характеристика этапов технологии KDD.
6. Data Mining. Характеристика классов задач, решаемых методами Data Mining.
7. Программный инструментарий для выполнения анализа данных.
8. Основные положения концепции хранилищ данных (DW).
9. Цели и задачи аффинитивного анализа. Поддержка и достоверность ассоциативных правил. Лифт и левередж.
10. Сферы применения ассоциативных правил.
11. Иерархические ассоциативные правила.
12. Цели, задачи и основное содержание кластерного анализа. Классификация методов кластеризации.
13. Способы определения меры расстояния между кластерами.
14. Характеристика методов связи для процедуры кластеризации (одиночная, полная, средняя).
15. Алгоритм кластеризации k-means.
16. Сети Кохонена (KCN).



- 17.Карты Кохонена (SOM).
- 18.Проблемы алгоритмов кластеризации.
- 19.Цели, задачи и отличительные особенности классификации и регрессии.
- 20.Сферы применения методов классификации и регрессии.
- 21.Простая линейная регрессия.
- 22.Оценка соответствия простой линейной регрессии реальным данным.
- 23.Простая регрессионная модель.
- 24.Оценка значимости простой регрессионной модели (t-критерий и F-критерий).
- 25.Множественная линейная регрессия.
- 26.Модель множественной линейной регрессии.
- 27.Оценка значимости множественной регрессионной модели.
- 28.Регрессия с категориальными входными переменными.
- 29.Методы отбора переменных в регрессионные модели.
- 30.Ограничения применимости регрессионных моделей.
- 31.Логистическая регрессия. Интерпретация модели логистической регрессии.
- 32.Множественная логистическая регрессия.
- 33.Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.
- 34.Сферы применения деревьев решений.
- 35.Алгоритмы ID3 и C4.5.
- 36.Алгоритм CART.
- 37.Упрощение деревьев решений.
- 38.Цели, задачи и принципы работы нейронных сетей.
- 39.Принципы функционирования многослойного персептрона.
- 40.Алгоритмы обучения нейронных сетей.
- 41.Алгоритм обратного распространения ошибки.
- 42.Общая характеристика временных рядов и их компонентов. Цели и задачи анализа временных рядов.
- 43.Цели, задачи и принципы прогнозирования. Модели прогнозирования. Обобщенная модель прогноза.
- 44.Ансамбли моделей. Бэггинг. Бустинг.
- 45.Альтернативные методы построения ансамблей.
- 46.Оценка эффективности и сравнение моделей.
- 47.Lift- и Profit-кривые.
- 48.ROC-анализ.

#### **7.4 Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.**

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля - 50% и промежуточного контроля - 50%.

Текущий контроль по дисциплине включает:

- посещение занятий- 10 баллов (на каждом занятии),
- участие на практических занятиях 90 - баллов (на каждом занятии),
- выполнение самостоятельной работы (написание, защита, презентация докладов, рефератов, выполнение лабораторных работи др.) - 100баллов.

Промежуточный контроль по дисциплине включает:

- письменная контрольная работа - 100 баллов, в т.ч.: теоретический вопрос - 20 баллов, тестирование – 30 баллов, решение задачи – 50 баллов.

#### **8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.**

##### ***а) основная литература:***

1. Айзек М.П. Вычисления, графики и анализ данных в Excel 2013 [Электронный ресурс]: самоучитель / М.П. Айзек, М.В. Финков, Р.Г. Прокди. — Электрон. текстовые данные. — СПб. : Наука и Техника, 2015. — 416 с. — Режим доступа: <http://www.iprbookshop.ru/35584.html> (1.09.18).
2. Дубров, Абрам Моисеевич. Многомерные статистические методы для экономистов и менеджеров : учебник / Дубров, Абрам Моисеевич ; В.С.Мхитарян, Л.И.Трошин. - М. : Финансы и статистика, 2005. - 350 с. : ил. - Рекомендовано МО. - ISBN 5-279-01945-3 : 121-00
3. Жуковский О.И. Информационные технологии и анализ данных [Электронный ресурс]: учебное пособие / О.И. Жуковский. — Электрон. текстовые данные. — Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. — 130 с. — Режим доступа: <http://www.iprbookshop.ru/72106.html> (1.09.18).
4. Хайкин, Саймон . Нейронные сети: полный курс : [пер. с англ.] / Хайкин, Саймон . - 2-е изд., испр. - М. : Вильямс, 2006. - 1103 с. : ил. - ISBN 5-8459-0890-6 : 711-75.

##### ***б) дополнительная литература:***

1. Галушкин, Александр Иванович. Нейронные сети: основы теории : [монография] / Галушкин, Александр Иванович. - М. : Горячая линия-Телеком, 2010. - 496 с. - Библиогр.: с. 469-488. - ISBN 978-5-9912-0082-0 : 418-00
2. Карабутов Н.Н. Создание интегрированных документов в Microsoft office. Введение в анализ данных и подготовку документов [Электронный ресурс] / Н.Н. Карабутов. — Электрон. текстовые данные. — М. : СОЛОН-ПРЕСС, 2010. — 293 с. — Режим доступа: <http://www.iprbookshop.ru/65138.html> (1.09.18).
3. Нейронные сети. STATISTICA Neural Networks : Пер. с англ. / StatSoft

- Russia. - М. : Горячая линия-Телеком: Грааль, 2000. - 181,[1] с. : ил. - ISBN 5-93517-015-9 : 0-0.
4. Тюрин, Юрий Николаевич. Анализ данных на компьютере / Тюрин, Юрий Николаевич, А. А. Макаров ; под ред. В.Э.Фигурнова . - 3-е изд., перераб. и доп. - М. : ИНФРА-М, 2003. - 543,[1] с. : ил. ; 22 см. - Библиогр.: с. 527-533. - ISBN 5-16-001316-4 : 200-00.
  5. Федин Ф.О. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс] : учебное пособие / Ф.О. Федин, Ф.Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 204 с. — Режим доступа: <http://www.iprbookshop.ru/26444.html> (1.09.18).
  6. Федин Ф.О. Анализ данных. Часть 2. Инструменты Data Mining [Электронный ресурс] : учебное пособие / Ф.О. Федин, Ф.Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 308 с. — Режим доступа: <http://www.iprbookshop.ru/26445.html> (1.09.18).

## **9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.**

<http://www.edu.basegroup.ru>

1. <http://www.olap.ru>
2. <http://basegroup.ru>
3. <http://www.gpss.ru>
4. <http://www.iisys.ru/>
5. <http://www.itstan.ru/it-i-is/ekspertnye-intellektualnye-is-informacionnye-sistemy.html>
6. <http://www.gorod-nevy.ru/>
7. <http://www.intelligentsystems.dk/en/methods.php>
8. <http://www.corportal.ru/Resources/DataTech/OLAP/>
9. <http://www.raai.org/>
10. <http://alicebot.blogspot.com/>
11. <http://www.olap.ru/desc/ibm/ibm.asp>
12. [http://www.itsec.ru/articles2/calendar/budusch\\_za\\_intellect\\_sist](http://www.itsec.ru/articles2/calendar/budusch_za_intellect_sist)

## **10. Методические указания для обучающихся по освоению дисциплины.**

Оптимальным путем освоения дисциплины является посещение всех лекций и семинаров и выполнение предлагаемых заданий в виде рефератов, докладов решения задач и устных вопросов.

На лекциях рекомендуется деятельность студента в форме активного слушания, т.е. предполагается возможность задавать вопросы на уточнение понимания темы и рекомендуется конспектирование основных положений

лекции. На практических занятиях деятельность студента заключается в активном обсуждении вопросов темы, докладов, рефератов, решении ситуационных задач, выполнении контрольных заданий и т.п.

При подготовке к практическому занятию студенты должны изучить конспект лекций по заданной теме, ознакомиться с соответствующим разделом в учебнике, рекомендованном в качестве основной литературы. Студент может ознакомиться и с дополнительной литературой: периодические издания, интернет-источники.

Форма работы с литературой может быть разнообразной – начиная от комментированного чтения и кончая выполнением различных заданий на основе прочитанной литературы. Например; составление плана, подбор выписок из литературы по заданным вопросам; конспектирование текста.

Подготовка к экзамену предполагает изучение конспектов лекций, рекомендуемой литературы, повторение материалов практических занятий

## **11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.**

### ***в) программное обеспечение:***

SQLServer 2008R2 AnalysisServices, аналитическая платформа DeductorStudio, пакет STATISTICA.

### ***г) базы данных, информационно-справочные и поисковые системы:***

образец базы данных OLTP AdventureWorks; образец хранилища данных AdventureWorksDW; образец базы данных службы аналитики AdventureWorksAS.

## **12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.**

Для проведения лекционных занятий: требуется аудитория, оборудованная меловой доской, интерактивной доской, мультимедийным проектором с экраном.

Для проведения лабораторных работ: требуется специализированный компьютерный класс.