

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

УТВЕРЖДАЮ  
Декан факультета  
/Фионова Л.Р./  
« 15 » июня 2015 г.



## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

### Б1.2.24.1 Методы интеллектуального анализа данных

Направление подготовки 01.03.02 Прикладная математика и информатика

Профиль подготовки Системное программирование и компьютерные технологии

Квалификация (степень) выпускника Бакалавр

Форма обучения очная

Пенза, 2015

## 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины "Методы интеллектуального анализа данных" является формирование и развитие у будущих бакалавров прикладной математики и информатики общепрофессиональных и профессиональных компетенций, формирование системы знаний, умений и навыков базовых методов интеллектуального анализа данных.

**Задачи изучаемой дисциплины.** Исходя из общих целей подготовки бакалавра прикладной математики и информатики по профилю "Системное программирование и компьютерные технологии":

- развитие способности использовать углубленные теоретические и практические знания в области прикладной математики и информатики;
- развитие способности понимать и применять в исследовательской и прикладной деятельности современный математический аппарат.

Исходя из специфики конкретной дисциплины:

- изучение основ построения систем поддержки принятия решений;
- освоение основных методов интеллектуального анализа данных;
- получение практических навыков работы в одной из современных аналитических платформ.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП БАКАЛАВРИАТА

Дисциплина "Методы интеллектуального анализа данных" относится к вариативной части блока дисциплин.

Для освоения дисциплины "Методы интеллектуального анализа данных" обучающиеся используют знания, умения и навыки, сформированные в ходе изучения дисциплин базовой части блока дисциплин: "Алгоритмы и алгоритмические языки", "Языки и методы программирования", "Базы данных и системы управления базами данных", "Численные методы" и дисциплин вариативной части: "Методы оптимизации" и "Нейронные сети".

Освоение данной дисциплины является также основой для последующего прохождения производственной и учебной практик, подготовки к итоговой государственной аттестации.

### 3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ "МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ"

Процесс изучения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО по данному направлению:

Коды компетенции	Наименование компетенции	Структурные элементы компетенции (в результате освоения дисциплины обучающийся должен знать, уметь, владеть)
ОПК-1	способность использовать базовые знания естественных наук, математики и информатики, основные факты, концепции, принципы теорий, связанных с прикладной математикой и информатикой	Знать: математические основы построения систем аналитической обработки данных
		Уметь: использовать методы интеллектуального анализа данных
		Владеть: основными методами интеллектуального анализа данных
ОПК-3	способность к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям	Знать: основные алгоритмические решения в области интеллектуального анализа данных
		Уметь: разрабатывать алгоритмические решения в области интеллектуального анализа данных
		Владеть: основными методами интеллектуального анализа данных
ОПК-4	способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Знать: основные стандартные задачи интеллектуального анализа данных
		Уметь: применять стандартные задачи интеллектуального анализа данных
		Владеть: методами решения стандартных задач интеллектуального анализа данных
ПК-7	способность к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения	Знать: основные программные решения в области интеллектуального анализа данных
		Уметь: применять программные решения в области интеллектуального анализа данных
		Владеть: основами работы в одной из современных аналитических платформ

## 4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ "МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ"

### 4.1. Структура дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов.

№ п/п	Наименование разделов и тем дисциплины (модуля)	Семестр	Недели семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)									Формы текущего контроля успеваемости (по неделям семестра)		
				Аудиторная работа			Самостоятельная работа								
				Всего	Лекции	Лабораторные занятия	Всего	Подготовка к лекциям	Выполнение индивидуальных домашних заданий	Подготовка к лабораторным работам	Подготовка к экзамену	Опрос на лабораторных занятиях	Проверка отчетов о выполнении лабораторных работ	Проверка индивидуальных домашних заданий	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
<b>1</b>	<b>РАЗДЕЛ 1. OLAP-СИСТЕМЫ</b>	<b>7</b>	<b>1–3</b>	<b>12</b>	<b>6</b>	<b>6</b>	<b>12</b>	<b>3</b>	<b>6</b>	<b>3</b>					
1.1	Тема 1.1. Основные принципы анализа данных	7	1	2	2		1	1				1			
1.2	Лабораторная работа 1. Освоение среды Deductor	7	1	2		2	3		2	1					
1.3	Тема 1.2. Концепция хранилища данных	7	2	2	2		1	1				2			
1.4	Лабораторная работа 1. Освоение среды Deductor	7	2	2		2	3		2	1			2	2	
1.5	Тема 1.3. OLAP-системы	7	3	2	2		1	1							
1.6	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	7	3	2		2	3		2	1					3
<b>2</b>	<b>РАЗДЕЛ 2. ПРОЦЕСС ETL — ИЗВЛЕЧЕНИЕ ДАННЫХ, ПРЕОБРАЗОВАНИЕ ДАННЫХ, ЗАГРУЗКА ДАННЫХ</b>	<b>7</b>	<b>4–5</b>	<b>8</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>2</b>	<b>4</b>	<b>2</b>					
2.1	Тема 2.1. Извлечение и преобразование данных	7	4	2	2		1	1				3			

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2.2	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	7	4	2		2	3		2	1			3	4
2.3	Тема 2.2. Очистка и предобработка данных	7	5	2	2		1	1				4		
2.4	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	7	5	2		2	3		2	1			4	5
<b>3</b>	<b>РАЗДЕЛ 3. АССОЦИАТИВНЫЕ ПРАВИЛА</b>	<b>7</b>	<b>6-7</b>	<b>8</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>2</b>	<b>4</b>	<b>2</b>				
3.1	Тема 3.1. Основы ассоциативных правил	7	6	2	2		1	1				5		
3.2	Лабораторная работа 3. Предобработка данных	7	6	2		2	3		2	1			5	6
3.3	Тема 3.2. Алгоритмы построения ассоциативных правил	7	7	2	2		1	1				6		
3.4	Лабораторная работа 4. Поиск ассоциативных правил	7	7	2		2	3		2	1			6	7
<b>4</b>	<b>РАЗДЕЛ 4. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ</b>	<b>7</b>	<b>8-14</b>	<b>28</b>	<b>14</b>	<b>14</b>	<b>28</b>	<b>7</b>	<b>14</b>	<b>7</b>				
4.1	Тема 4.1. Логистическая регрессия	7	8	2	2		1	1						
4.2	Лабораторная работа 5. Классификация с помощью логистической регрессии	7	8	2		2	3		2	1				8
4.3	Тема 4.2. Оценка и сравнение классификаторов. ROC-анализ	7	9	2	2		1	1						
4.4	Лабораторная работа 6. Оценка качества классификаторов	7	9	2		2	3		2	1				9
4.5	Тема 4.3. Правила классификации. Алгоритм "1 правило"	7	10	2	2		1	1						
4.6	Лабораторная работа 7. Классификация с помощью деревьев решений	7	10	2		2	3		2	1				10
4.7	Тема 4.4. Байесовская классификация	7	11	2	2		1	1						
4.8	Лабораторная работа 7. Классификация с помощью деревьев решений	7	11	2		2	3		2	1				11
4.9	Тема 4.5. Деревья решений. Алгоритмы ID3	7	12	2	2		1	1				8		
4.10	Лабораторная работа 8. Классификация с помощью нейронных сетей	7	12	2		2	3		2	1			8	12
4.11	Тема 4.6. Деревья решений. Алгоритм C4.5 и CART	7	13	2	2		1	1				9		
4.12	Лабораторная работа 8. Классификация с помощью нейронных сетей	7	13	2		2	3		2	1			9	13
4.13	Тема 4.7. Решение задач классификации и регрессии с помощью нейронных сетей	7	14	2	2		1	1				10		
4.14	Лабораторная работа 9. Прогнозирование с помощью нейронных сетей	7	14	2		2	3		2	1			10	14

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<b>5</b>	<b>РАЗДЕЛ 5. КЛАСТЕРИЗАЦИЯ</b>	<b>7</b>	<b>15–16</b>	<b>8</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>2</b>	<b>4</b>	<b>2</b>				
5.1	Тема 5.1. Алгоритмы кластеризации	7	15	2	2		1	1				11		
5.2	Лабораторная работа 10. Кластеризация с помощью алгоритма k-means	7	15	2		2	3		2	1			11	15
5.3	Тема 5.2. Кластеризация с использованием сетей Кохонена	7	16	2	2		1	1				12		
5.4	Лабораторная работа 11. Кластеризация с помощью самоорганизующихся карт Кохонена	7	16	2		2	3		2	1			12	16
<b>6</b>	<b>РАЗДЕЛ 6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ</b>	<b>7</b>	<b>17–18</b>	<b>8</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>2</b>	<b>4</b>	<b>2</b>				
6.1	Тема 6.1. Анализ временных рядов	7	17	2	2		1	1				13		
6.2	Лабораторная работа 12. Анализ временных рядов	7	17	2		2	3		2	1			13	17
6.3	Тема 6.2. Прогнозирование временных рядов	7	18	2	2		1	1				14		
6.4	Лабораторная работа. Итоговое занятие	7	18	2		2	3		2	1			14	18
	<b>Подготовка к экзамену</b>						<b>36</b>				<b>36</b>			
	<b>Общая трудоемкость, в часах</b>			<b>72</b>	<b>36</b>	<b>36</b>	<b>108</b>	<b>18</b>	<b>36</b>	<b>18</b>	<b>36</b>	Промежуточная аттестация		
												Форма	Се- местр	
													Экзамен	7

## **4.2. Содержание дисциплины**

### **4.2.1. Содержание лекционных занятий**

#### **РАЗДЕЛ 1. OLAP-СИСТЕМЫ**

##### ***Тема 1.1. Основные принципы анализа данных***

Понятие системы поддержки принятия решений. OLAP — оперативная аналитическая обработка данных. KDD — извлечение знаний из баз данных. Этапы KDD. Data Mining — "добыча данных". Основные задачи Data Mining.

##### ***Тема 1.2. Концепция хранилища данных***

Требования, предъявляемые к базам данных в СППР. Концепция хранилища данных. Архитектура СППР. Многомерная модель данных.

##### ***Тема 1.3. OLAP-системы***

Понятие OLAP-системы, правила Кодда, тест FASMI. Реализация OLAP-систем: MOLAP, ROLAP, HOLAP.

#### **РАЗДЕЛ 2. ПРОЦЕСС ETL — ИЗВЛЕЧЕНИЕ ДАННЫХ, ПРЕОБРАЗОВАНИЕ ДАННЫХ, ЗАГРУЗКА ДАННЫХ**

##### ***Тема 2.1. Извлечение и преобразование данных***

Составляющие процесса ETL. Извлечение данных в ETL. Трансформация данных: преобразование временных данных, группировка и слияние, квантование, нормализация.

##### ***Тема 2.2. Очистка и предобработка данных***

Предобработка данных. Инструменты предобработки в аналитическом приложении. Обработка дубликатов, противоречий и аномалий. Восстановление пропущенных данных. Сокращение размерности исходных данных. Сэмплинг.

#### **РАЗДЕЛ 3. АССОЦИАТИВНЫЕ ПРАВИЛА**

##### ***Тема 3.1. Основы ассоциативных правил***

Понятие машинного обучения. Понятие ассоциативного правила. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил: лифт, левередж.

##### ***Тема 3.2. Алгоритмы построения ассоциативных правил***

Алгоритм Apriori: частые предметные наборы и их обнаружение, генерация ассоциативных правил. Иерархические ассоциативные правила. Последовательные шаблоны.

#### **РАЗДЕЛ 4. КЛАССИФИКАЦИЯ И РЕГРЕССИЯ**

##### ***Тема 4.1. Логистическая регрессия***

Понятие логистической регрессии. Нахождение коэффициентов логистической регрессии методом максимального правдоподобия. Аналогия с искусственным нейроном.

##### ***Тема 4.2. Оценка и сравнение классификаторов. ROC-анализ***

Ошибки классификации. Матрица несоответствий. Ошибки первого и второго рода. Чувствительность и специфичность. Точка отсечения. ROC-кривая.

##### ***Тема 4.3. Правила классификации. Алгоритм "1 правило"***

Правила классификации. Алгоритм "1 правило" построения правил классификации.

##### ***Тема 4.4. Байесовская классификация***

Формула Байеса. Классификация на основе формулы Байеса. Алгоритм простого (наивного) байесовского классификатора.

##### ***Тема 4.5. Деревья решений. Алгоритм ID3***

Понятие дерева решений. "Жадные" алгоритмы построения деревьев решений. Выбор атрибута ветвления: индекс Джини, критерий уменьшения энтропии, критерий отношения прироста информации. Алгоритмы ID3.

#### ***Тема 4.6. Деревья решений. Алгоритмы C4.5 и CART***

Проблемы переобучения и неизвестных значений атрибутов. Алгоритм C4.5. Алгоритм CART. Регрессионное дерево решений. Принципы упрощения деревьев решений.

#### ***Тема 4.7. Решение задач классификации и регрессии с помощью нейронных сетей***

Особенности использования нейронных сетей для решения задач классификации и регрессии.

### **РАЗДЕЛ 5. КЛАСТЕРИЗАЦИЯ**

#### ***Тема 5.1. Алгоритмы кластеризации***

Постановка задачи кластеризации. Меры близости, используемые в кластеризации. Иерархические алгоритмы кластеризации. Неиерархические алгоритмы кластеризации.

#### ***Тема 5.2. Кластеризация с использованием сетей Кохонена***

Обучение сети Кохонена. Карты Кохонена: построение карты, выбор числа нейронов.

### **РАЗДЕЛ 6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ**

#### ***Тема 6.1. Анализ временных рядов***

Понятие временного ряда. Компоненты ряда. Модели ряда. Корреляционный анализ временных рядов.

#### ***Тема 6.2. Прогнозирование временных рядов***

Понятие прогнозирования временного ряда. Прогнозирование методом среднего и скользящего среднего. Регрессионные модели прогнозирования. Моделирование временного ряда случайным процессом.

#### **4.2.2. Темы лабораторных работ**

1. Освоение среды Deductor.
2. Проектирование хранилищ данных Deductor Warehouse.
3. Предобработка данных.
4. Поиск ассоциативных правил.
5. Классификация с помощью логистической регрессии.
6. Оценка качества классификаторов
7. Классификация с помощью деревьев решений.
8. Классификация с помощью нейронных сетей.
9. Прогнозирование с помощью нейронных сетей.
10. Кластеризация с помощью алгоритма k-means.
11. Кластеризация с помощью самоорганизующихся карт Кохонена.
12. Анализ временных рядов.

### **5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ**

В ходе освоения дисциплины "Методы интеллектуального анализа данных" при проведении аудиторных занятий используется образовательная технология, предусматривающая такие методы и формы изучения материала как лекция и лабораторное занятие, включающие активные и интерактивные формы занятий:



Проведение лекции проблемного характера: тема 2.2. "Очистка и преобработка данных"; тема 4.5. "Деревья решений. Алгоритмы ID3 и C4.5; тема 5.2. "Кластеризация с использованием сетей Кохонена".

Проведение лабораторных занятий в интерактивной форме и публичная защита отчетов по лабораторным работам, работа в малых группах: лабораторная работа 2 "Проектирование хранилищ данных Deductor Warehouse "; лабораторная работа 9 "Прогнозирование с помощью нейронных сетей"; лабораторная работа 10 "Кластеризация с помощью самоорганизующихся карт Кохонена".

Занятия, проводимые в интерактивной форме, составляют 25 % от общего количества аудиторных занятий.

Лабораторные занятия проводятся с использованием свободно распространяемой современной аналитической платформы Deductor Academic. При выполнении лабораторных работ и индивидуальных домашних заданий используются исходные данные из репозитория UCI Калифорнийского университета в Ирвайне (University of California, Irvine — <http://archive.ics.uci.edu/ml/datasets.html>) и тестовые данные фирмы BaseGroup Labs — [www.basegroup.ru](http://www.basegroup.ru)).

Самостоятельная работа студентов подразумевает работу под руководством преподавателя (консультации, помощь в написании и отладке программ и др.) и индивидуальную работу студента, выполняемую как дома, так и в компьютерном классе с выходом в Интернет.

При реализации образовательных технологий используются следующие виды самостоятельной работы:

- работа с конспектом лекции и литературой;
- подготовка к лабораторной работе: изучение теоретического материала, разработка и отладка программ заданий по лабораторным работам;
- обработка результатов лабораторных работ и подготовка письменных отчетов;
- выполнение и оформление индивидуальных домашних заданий: изучение теоретического материала, разработка алгоритма решения задачи, разработка и отладка программ, эксперимент с разработанной программой, оформление письменного отчета;
- поиск информации в Интернет и литературе;
- подготовка к сдаче лабораторных работ и индивидуальных заданий;
- подготовка к сдаче зачёта.

В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по собственной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины базируется на следующих возможностях: обеспечение внеаудиторной работы со студентами в том числе в электронной образовательной среде с использованием соответствующего программного оборудования, дистанционных форм обучения, возможностей интернет-ресурсов, индивидуальных консультаций и т.д.

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.

### 6.1. План самостоятельной работы студентов

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
1	Тема 1.1. Основные принципы анализа данных	Подготовка к лекциям	Изучить основные понятия систем поддержки принятия решений и OLAP, Основные задачи Data Mining. Найти в Internet, например, на сайте <a href="http://www.basegroup.ru/">http://www.basegroup.ru/</a> примеры использования Data Mining.	1–5	1
1	Лабораторная работа 1. Освоение среды Deductor	Подготовка к лабораторным работам	Изучить рабочую среду Deductor Academic.	4	1
		Выполнение индивидуального задания	Анализ задания, подбор литературы.	1–4	2
2	Тема 1.2. Концепция хранилища данных	Подготовка к лекциям	Изучить требования, предъявляемые к базам данных в СППР, концепцию хранилища данных, различные реализации OLAP-систем. По документации Deductor Academic узнать, как реализуется хранилище данных в этой системе, какие возможности в плане работы с хранилищем имеют академическая и коммерческая версии продукта.	1, 2, 4	1
2	Лабораторная работа 1. Освоение среды Deductor	Подготовка к лабораторным работам	Изучить базовые визуализаторы и основные возможности Data Mining платформы Deductor Academic. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Анализ задания, подбор литературы.	1–5	2
3	Тема 1.3. OLAP-системы	Подготовка к лекциям	Изучить понятие, основные свойства и реализации OLAP-систем. По Internet ознакомиться с основными коммерческими OLAP-системами.	1, 4	1
3	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	Подготовка к лабораторным работам	Изучить принципы построения хранилища данных Deductor Warehouse. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Загрузка исходных данных, перевод на русский язык описания исходных данных.	4	2
4	Тема 2.1. Извлечение и преобразование данных	Подготовка к лекциям	Изучить виды трансформации данных: преобразование временных данных, группировка и слияние, квантование, нормализация.	1, 4	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
4	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	Подготовка к лабораторным работам	Разработать структуру хранилища данных. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Изучить методы подготовки данных в Deductor.	4	2
5	Тема 2.2. Очистка и предобработка данных	Подготовка к лекциям	Изучить основные операции предобработки данных: обработка дубликатов, противоречий и аномалий, восстановление пропущенных данных, сокращение размерности исходных данных. Сравнить изучаемые методы предобработки данных с методами, изученными в дисциплине "Нейронные сети".	1, 4	1
5	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	Подготовка к лабораторным работам	Освоить работу с хранилищем данных Deductor Academic. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Оценка качества исходных данных, подготовка данных в Deductor.	4	2
6	Тема 3.1. Основы ассоциативных правил	Подготовка к лекциям	Изучить понятие ассоциативного правила и числовые характеристики правила: поддержка, достоверность, лифт, леввередж.	1, 4	1
6	Лабораторная работа 3. Предобработка данных	Подготовка к лабораторным работам	Изучить алгоритмы предварительной обработки данных в Deductor Academic. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Изучение методов анализа данных в Deductor.	4	2
7	Тема 3.2. Алгоритмы построения ассоциативных правил	Подготовка к лекциям	Изучить алгоритм Apriori.	1, 4	1
7	Лабораторная работа 4. Поиск ассоциативных правил	Подготовка к лабораторным работам	Изучить поиск ассоциативных правил в Deductor Academic. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Изучение методов анализа данных в Deductor.	4	2
8	Тема 4.1. Логистическая регрессия	Подготовка к лекциям	Изучить логистическую регрессию. Сравнить логистическую регрессию с нейроном. В чем преимущества нейронной сети перед логистической регрессией?	1, 3, 4, 5	1
8	Лабораторная работа 5. Классификация с помощью логистической регрессии	Подготовка к лабораторным работам	Изучить логистическую регрессию. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Выбор методов анализа для конкретного задания.	4	2
9	Тема 4.2. Оценка и сравнение классификаторов.	Подготовка к лекциям	Изучить сравнение классификаторов и ROC-анализ.	1, 3, 4, 5	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
	ROC-анализ		Повторить понятия ошибок первого и второго рода.		
9	Лабораторная работа 6. Оценка качества классификаторов	Подготовка к лабораторным работам	На примере логистической регрессии изучить оценки качества классификаторов Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Доработка перечня методов анализа с учетом замечаний преподавателя.	4	2
10	Тема 4.3. Правила классификации. Алгоритм "1 правило"	Подготовка к лекциям	Изучить алгоритм "1 правило" построения правил классификации.	1, 4	1
10	Лабораторная работа 7. Классификация с помощью деревьев решений	Подготовка к лабораторным работам	Изучить деревья классификации в Deductor Academic. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	1, 3, 4, 5	2
11	Тема 4.4. Байесовская классификация	Подготовка к лекциям	Повторить теорему Байеса. Изучить байесовскую классификацию, "наивный"-байесовский подход.	1, 3, 4, 5	1
11	Лабораторная работа 7. Классификация с помощью деревьев решений	Подготовка к лабораторным работам	Изучить построение интерактивных деревьев решений в Deductor Academic. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
12	Тема 4.5. Деревья решений. Алгоритмы ID3	Подготовка к лекциям	Изучить методы выбора атрибута ветвления и алгоритм ID3 построения дерева решения.	1, 3, 4, 5	1
12	Лабораторная работа 8. Классификация с помощью нейронных сетей	Подготовка к лабораторным работам	Повторить структуру и принципы обучения многослойного перцептрона. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	1, 3, 4, 5	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
13	Тема 4.6. Деревья решений. Алгоритм C4.5 и CART	Подготовка к лекциям	Изучить алгоритмы C4.5 и CART, понятие регрессионного дерева решений, принципы упрощения деревьев решений.	4	1
13	Лабораторная работа 8. Классификация с помощью нейронных сетей	Подготовка к лабораторным работам	Изучить реализацию нейросетевой классификации в Deductor Academic. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
14	Тема 4.7. Решение	Подготовка к лекциям	Изучить взаимодействие нейронных сетей с другими методами Data Mining. Изучить	1, 3, 4, 5	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
	задач классификации и регрессии с помощью нейронных сетей		использование нейронных сетей для решения задач классификации и регрессии.		
14	Лабораторная работа 9. Прогнозирование с помощью нейронных сетей	Подготовка к лабораторным работам	Изучить реализацию нейросетевого прогнозирования в Deductor Academic. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
15	Тема 5.1. Алгоритмы кластеризации	Подготовка к лекциям	Изучить меры близости, используемые в кластеризации, иерархические и неиерархические алгоритмы кластеризации.	1, 3, 5, 5	1
15	Лабораторная работа 10. Кластеризация с помощью алгоритма k-means	Подготовка к лабораторным работам	Изучить реализацию алгоритма k-means. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Оформление отчета о выполнении индивидуального задания.		2
16	Тема 5.2. Кластеризация с использованием сетей Кохонена	Подготовка к лекциям	Изучить построение сетей и карт Кохонена. Повторить обучение сетей Кохонена.	4	1
16	Лабораторная работа 11. Кластеризация с помощью самоорганизующихся карт Кохонена	Подготовка к лабораторным работам	Изучить реализацию в Deductor Academic алгоритма кластеризации с помощью карт Кохонена. Подготовка к защите лабораторной работы.	4	2
		Выполнение индивидуального задания	Оформление отчета о выполнении индивидуального задания. Подготовка к защите отчета.	1, 3, 4, 5	2
17	Тема 6.1. Анализ временных рядов	Подготовка к лекциям	<i>Работа с конспектом лекции и литературой</i> Изучить понятие и модели временного ряда. Изучить анализ временных рядов.	1	1
17	Лабораторная работа 12. Анализ временных рядов	Подготовка к лабораторным работам	Изучить анализ временных рядов. Оформление отчета о лабораторной работе.	1, 3, 4, 5	2
		Выполнение индивидуального задания	Защита отчета.		2
18	Тема 6.2. Прогнозирование временных рядов	Подготовка к лекциям	Изучить прогнозирование временного ряда методами среднего и скользящего среднего. Изучить регрессионные модели прогнозирования и моделирование временного ряда случайным процессом.	1, 3, 4, 5	1
18	Лабораторная работа. Итоговое занятие	Подготовка к лабораторным работам	Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	1, 3, 4, 5	1
		Выполнение индивидуального задания	Защита отчета.	1, 3, 4, 5	2

## 6.2. Методические указания по организации самостоятельной работы студентов

При работе с конспектом лекций и изучении рекомендованной литературы студенту необходимо изучить конспект лекций, ответить на контрольные вопросы, изучить разделы рекомендованной литературы. Следует поощрять регулярную работу студентов с теоретическим материалом и чтение источников, выходящих за пределы рекомендованного списка литературы.

При подготовке к лабораторным работам студентам следует изучить теоретический материал по теме лабораторной работы, ответить на контрольные вопросы

При оформлении отчетов по лабораторным работам студент должен изучить требования к оформлению отчета, представить результаты выполнения работы, проанализировать результаты работы и сделать выводы по работе.

При выполнении индивидуального задания студенту необходимо провести анализ задания, изучить рекомендованную литературу, обоснованно выбрать метод решения задач, разработать алгоритм решения, провести эксперименты, используя аналитическую платформу Deductor Academic, проанализировать результаты, сравнить различные аналитические методы.

Подготовка к экзамену подразумевает повторение изученного материала. Использование при подготовке и ответах результатов выполнения индивидуальных заданий облегчает подготовку и повышает качество ответа.

Студентам из числа лиц с ограниченными возможностями здоровья могут быть предложены электронные образовательные ресурсы в формах, адаптированных к ограничениям их здоровья.

## 6.3. Материалы для проведения текущего и промежуточного контроля знаний студентов

№ п/п	Вид контроля	Контролируемые разделы	Компетенции, компоненты которых контролируются
1.	Текущий: опрос на лабораторных работах.	1–6	ОПК-1, 3, 4, ПК-7
2.	Проверка отчетов о выполнении лабораторных работ.	1–6	ОПК-1, 3, 4, ПК-7
3.	Проверка индивидуальных домашних заданий	2, 4–5	ОПК-1, 3, 4, ПК-7
4.	Промежуточный: экзамен	1–6	ОПК-1, 3, 4, ПК-7

### Демонстрационные варианты индивидуальных домашних заданий

С использованием платформы Deductor Academic провести диагностику рака молочной железы — классификацию опухоли как доброкачественной или злокачественной в зависимости от характеристик образцов биопсии. Для этого, используя исходные данные из репозитория UCI (Breast Cancer — <http://archive.ics.uci.edu/ml/datasets/Breast+Cancer>), сформировать

базу данных для решения задачи. Произвести очистку данных. Выбрать факторы для построения модели. Исследовать возможности применения различных моделей (логистическая регрессия, нейронные сети, деревья решений, карты Кохонена). Дать анализ полученных результатов, подготовить отчет.

### Перечень примерных вопросов к экзамену

1. Понятие систем поддержки принятия решений. Knowledge Discovery in Databases и Data Mining.
2. Основные задачи, решаемые в Data Mining.
3. CRISP-DM — кросс-индустриальный стандарт Data Mining. Реализация Data Mining.
4. Требования, предъявляемые к базам данных в СППР.
5. Концепция хранилища данных.
6. Типы архитектур СППР.
7. Многомерная модель данных.
8. Понятие OLAP-системы.
9. Способы реализации хранилищ данных. Архитектура MOLAP.
10. Технологии ROLAP и HOLAP реализации хранилищ данных.
11. Процесс ETL.
12. Трансформация данных.
13. Инструменты предобработки в аналитическом приложении.
14. Обработка дубликатов, противоречий и аномалий.
15. Восстановление пропущенных данных.
16. Декорреляция входных данных.
17. Понижение размерности исходных данных.
18. Понятие сэмплинга.
19. Понятие машинного обучения.
20. Понятие ассоциативных правил. Поддержка, достоверность, лифт.
21. Алгоритм Apriori построения ассоциативных правил.
22. Иерархические ассоциативные правила.
23. Логистическая регрессия.
24. Оценка и сравнение классификаторов. ROC-анализ.
25. Алгоритм "1-правило" построения правил классификации.
26. Байесовская классификация. Простой ("наивный") байесовский классификатор.
27. Понятие дерева решений.
28. Алгоритм ID3 построения дерева решений.
29. Алгоритм C4.5 построения дерева решений.
30. Принципы упрощения деревьев решений.
31. Решение задач классификации и регрессии с помощью нейронных сетей
32. Постановка задачи кластеризации. Меры близости, используемые в кластеризации.
33. Иерархические методы кластеризации.
34. Неиерархические методы кластеризации. Адаптивные методы кластеризации.
35. Принципы построения сети Кохонена.
36. Обучение сети Кохонена.
37. Принципы построения карт Кохонена. Алгоритмы обучения карты Кохонена.
38. Понятие временного ряда.
39. Методы анализа временных рядов.
40. Методы прогнозирования временных рядов.

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ "МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ"

### 7.1. Основная литература

1. Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — СПб.: БХВ-Петербург, 2009. — 512 с.  
Электронная библиотека Попечительского совета механико-математического факультета Московского государственного университета <http://lib.mexmat.ru/books/28389>
2. Макарычев П. П., Механов В. Б., Афонин А. Ю. Оперативный и интеллектуальный анализ данных. — Пенза: Изд-во Пенз. гос. ун-та, 2010. — 156 с. (29 экз.)  
[http://kleopatra.pnzgu.ru/cgi-bin/irbis64r\\_91/cgiirbis\\_64.exe?P21DBN=KATL&I21DBN=KATL\\_PRINT&S21FMT=fullw\\_print&C21COM=F&Z21MFN=13693](http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=13693)
3. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. — М.: ДМК Пресс, 2015. — 400 с.  
(ЭБС "Лань". Режим доступа: <http://e.lanbook.com/book/69955>)

### 7.2. Дополнительная литература

5. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. — СПб.: Питер, 2013. — 704 с. (5 экз.)  
[http://kleopatra.pnzgu.ru/cgi-bin/irbis64r\\_91/cgiirbis\\_64.exe?P21DBN=KATL&I21DBN=KATL\\_PRINT&S21FMT=fullw\\_print&C21COM=F&Z21MFN=14890](http://kleopatra.pnzgu.ru/cgi-bin/irbis64r_91/cgiirbis_64.exe?P21DBN=KATL&I21DBN=KATL_PRINT&S21FMT=fullw_print&C21COM=F&Z21MFN=14890)
5. Рашка С. Python и машинное обучение. — М.: ДМК Пресс, 2017. — 418 с.  
(ЭБС "Лань". Режим доступа: <https://e.lanbook.com/book/100905>)

### 7.3. Интернет-ресурсы

В Интернет имеется огромное количество ресурсов, посвященных аналитике и системам поддержки принятия решений. В таблице перечислены наиболее авторитетные из них.

№ п/п	Адрес сайта	Описание материала, содержащегося на сайте
1.	<a href="http://www.olap.ru">www.olap.ru</a>	Российский сайт, посвященный Business Intelligence и OLAP-технологии. Содержит новости и статьи по указанной тематике.
2.	<a href="http://www.basegroup.ru">www.basegroup.ru</a>	Сайт фирмы BaseGroup Labs. Содержит много материалов по методологии Data Mining и системе Deductor. Можно скачать свободно распространяемую версию Deductor Academic.
3.	<a href="http://www.intuit.ru/department/database/datamining/">http://www.intuit.ru/department/database/datamining/</a>	Чубукова И. А. Курс: Data Mining
4.	<a href="http://www.intuit.ru/department/database/datawarehouse/">http://www.intuit.ru/department/database/datawarehouse/</a>	Перминов Г. И. Видеокурс: Хранилища данных
5.	<a href="http://www.intuit.ru/studies/courses/3498/740/info">http://www.intuit.ru/studies/courses/3498/740/info</a>	Видеокурс: Алгоритмы интеллектуальной обработки больших объемов данных.
6.	<a href="http://www.machinelearning.ru">http://www.machinelearning.ru</a>	MachineLearning.ru — профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
	<a href="https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie">https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie</a>	Интерактивный курс Введение в машинное обучение Школы анализа данных Яндекс:



	<a href="https://yandexdataschool.ru/edu-process/courses/machine-learning#item-1">https://yandexdataschool.ru/edu-process/courses/machine-learning#item-1</a>	Видеолекции "Машинное обучение" Школы анализа данных Яндекс.
	<a href="https://ru.coursera.org/specializations/machine-learning">https://ru.coursera.org/specializations/machine-learning</a>	Интерактивный курс на Coursera Build Intelligent Applications (на английском языке).

#### **7.4. Программное обеспечение**

Все лабораторные работы выполняются на персональных компьютерах на свободно распространяемой аналитической платформе Deductor Academic.

### **8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ "МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ"**

Студенты используют рабочие места в компьютерном классе, оборудованном локальной сетью и выходом в Internet, имеющиеся в библиотеке учебники. На сетевом ресурсе университета представлены в свободном доступе входящие в состав учебно-методического комплекса полные электронные версии лекционного курса, лабораторный практикум, индивидуальные задания и библиотека программ примеров (автор Горбаченко В. И.). Все лабораторные работы выполняются на персональных компьютерах.


Рабочая программа дисциплины "Методы интеллектуального анализа данных" составлена в соответствии с требованиями ФГОС ВО по направлению подготовки "Прикладная математика и информатика".

Программу составил  
д.т.н., профессор

 Горбаченко В. И.

**Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.**


Программа одобрена на заседании кафедры "Компьютерные технологии"  
протокол № 10 от "10" июня 20 15 года

Зав. каф. "Компьютерные технологии" д.т.н., профессор  Горбаченко В. И.

Программа одобрена методической комиссией факультета ВТ

протокол № 6 от «15» июня 20 15 г.

Председатель методической комиссии факультета ВТ

  
(подпись)

И.И. Копылов  
(Ф.И.О.)

**Сведения о переутверждении программы на очередной учебный год и регистрации изменений**

Учебный год	Решение кафедры (№ протокола, дата, подпись зав. кафедрой)	Внесенные изменения	Номера листов (страниц)		
			замененных	новых	аннулированных
2016/2017	№ 1 от 31.02.2016 ✓	изменен - № 5	с. 9	—	—
2017/2018	№ 1 от 30.02.2017 ✓	без изменений	—	—	—