

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

УТВЕРЖДАЮ
Декан факультета



[Signature] /Фионова Л.Р./

« 17 » *апреля* 2015 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.2.18 Методы интеллектуального анализа данных

Направление подготовки 01.03.02 Прикладная математика и информатика

Профиль подготовки Системное программирование и компьютерные технологии

Квалификация (степень) выпускника Академический 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 общекультурных 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 ООП 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 и нечеткие системы".

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-5	способность осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в информационно-телекоммуникационной сети "Интернет" (далее — сеть "Интернет") и в других источниках	Знать: основные источники информации по интеллектуальному анализу данных в сети Интернет
		Уметь: проводить поиск информации в сети Интернет
		Владеть: навыками поиска информации в сети Интернет
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 4 зачетных единицы, 144 часа.

№ 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						Опрос на лабораторных занятиях	Проверка отчетов о выполне- нии лабораторных работ	Проверка индивидуальных домашних заданий	Контрольные работы
				Всего	Лекции	Лабораторные занятия	Всего	Подготовка к лекциям	Выполнение индивидуальных домашних заданий	Подготовка к лабораторным работам	Подготовка к контрольным работам	Подготовка к экзамену				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	РАЗДЕЛ 1. OLAP-СИСТЕМЫ	7	1–2	8	4	4	4	2		2						
1.1	Тема 1.1. Основные принципы анализа данных	7	1	2	2		1	1					1			
1.2	Лабораторная работа 1. Освоение среды Deductor Studio	7	1	2		2	1			1						
1.3	Тема 1.2. Концепция хранилища данных. OLAP-системы	7	2	2	2		1	1					2			
1.4	Лабораторная работа 1. Освоение среды Deductor Studio	7	2	2		2	1			1				2		
2	РАЗДЕЛ 2. ПРОЦЕСС ETL — ИЗВЛЕЧЕНИЕ ДАННЫХ, ПРЕОБРАЗОВАНИЕ ДАННЫХ, ЗАГРУЗКА ДАННЫХ	7	3–4	8	4	4	6	2	2	2						
2.1	Тема 2.1. Извлечение и преобразование данных	7	3	2	2		1	1					3			
2.2	Лабораторная работа 2. Проектирование хранилищ дан- ных Deductor Warehouse	7	3	2		2	1			1				3		
2.3	Тема 2.2. Очистка и предобработка данных	7	4	2	2		1	1					4			

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2.4	Лабораторная работа 2. Проектирование хранилищ данных Deductor Warehouse	7	4	2		2	3		2	1				4	4	
3	РАЗДЕЛ 3. АССОЦИАТИВНЫЕ ПРАВИЛА	7	5–6	8	4	4	8	2	4	2						
3.1	Тема 3.1. Основы ассоциативных правил	7	5	2	2		1	1					5			
3.2	Лабораторная работа 3. Предобработка данных	7	5	2		2	3		2	1				5	5	
3.3	Тема 3.2. Алгоритмы построения ассоциативных правил	7	6	2	2		1	1					6			
3.4	Лабораторная работа 4. Поиск ассоциативных правил	7	6	2		2	3		2	1				6	6	
4	РАЗДЕЛ 4.	7	7-10	16	8	8	16	4	8	4						
4.1	Тема 4.1. Правила классификации	7	7	2	2		1	1					7			
4.2	Лабораторная работа 5. Прогнозирование с помощью логистической регрессии	7	7	2		2	3		2	1				7	7	
4.3	Тема 4.2. Деревья решений. Алгоритмы ID3 и C4.5.	7	8	2	2		1	1					8			
4.4	Лабораторная работа 6. Классификация с помощью деревьев решений	7	8	2		2	3		2	1				8	8	
4.5	Тема 4.3. Деревья решений. Алгоритм CART	7	9	2	2		1	1					9			
4.6	Лабораторная работа 7. Классификация с помощью нейронных сетей	7	9	2		2	3		2	1				9	9	
4.7	Тема 4.4. Решение задач классификации и регрессии с помощью нейронных сетей	7	10	2	2		1	1					10			
4.8	Лабораторная работа 8. Прогнозирование с помощью нейронных сетей	7	10	2		2	3		2	1				10	10	
5	РАЗДЕЛ 5. КЛАСТЕРИЗАЦИЯ	7	11-12	8	4	4	10	2	4	2	2					
5.1	Тема 5.1. Алгоритмы кластеризации	7	11	2	2		1	1					11			
5.2	Лабораторная работа 9. Кластеризация с помощью алгоритма k-means	7	11	2		2	5		2	1	2			11	11	11
5.3	Тема 5.2. Кластеризация с использованием сетей Кохонена	7	12	2	2		1	1					12			
5.4	Лабораторная работа 10. Кластеризация с помощью самоорганизующихся карт Кохонена	7	12	2		2	3		2	1				12	12	
6	РАЗДЕЛ 6. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ	7	13–14	8	4	4	8	2	4	2						
6.1	Тема 6.1. Анализ временных рядов	7	13	2	2		1	1					13			
6.2	Лабораторная работа 11. Корреляционный анализ временного ряда	7	13	2		2	3		2	1				13	13	

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17			
6.3	Тема 6.2. Прогнозирование временных рядов	7	14	2	2		1	1					14						
6.4	Лабораторная работа. Итоговое занятие	7	14	2		2	3		2	1				14	14				
	Общая трудоемкость, в часах			56	28	28	88	14	22	14	2	36	Промежуточная аттестация						
																	Форма	Семестр	
																	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. OLAP-системы

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

Правила классификации. Алгоритм "1 правило" построения правил классификации. Байесовская классификация, "наивно"-байесовский подход (Naive-Bayes approach).

Тема 4.2. Деревья решений. Алгоритмы ID3 и C4.5.

Понятие дерева решений. "Жадные" алгоритмы построения деревьев решений. Выбор атрибута ветвления: индекс Джини, критерий уменьшения энтропии, критерий отношения прироста информации. Алгоритмы ID3, проблемы переобучения и неизвестных значений атрибутов. Алгоритм C4.5.

Тема 4.3. Деревья решений. Алгоритм CART

Алгоритм CART. Регрессионное дерево решений. Принципы упрощения деревьев решений.

Тема 4.4. Решение задач классификации и регрессии с помощью нейронных сетей

Взаимодействие нейронных сетей с другими методами Data Mining. Использование нейронных сетей для решения задач классификации и регрессии. Нейронечеткие классификаторы.

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 Studio.
2. Проектирование хранилищ данных Deductor Warehouse.
3. Предобработка данных.
4. Поиск ассоциативных правил.
5. Прогнозирование с помощью логистической регрессии.
6. Классификация с помощью деревьев решений.
7. Классификация с помощью нейронных сетей.
8. Прогнозирование с помощью нейронных сетей.
9. Кластеризация с помощью алгоритма k-means.
10. Кластеризация с помощью самоорганизующихся карт Кохонена.
11. Корреляционный анализ временного ряда.

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, лабораторное занятие, включающие активные и интерактивные формы занятий:

Проведение лекции проблемного характера: тема 2.2. "Очистка и предобработка данных"; тема 4.2. "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 8 "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 в современной аналитической платформе Deductor. При выполнении лабораторных работ и индивидуальных домашних заданий используются исходные данные из репозитория UC1 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).

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 контрольным работам;
- 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	1.1	Подготовка к лекциям	Изучить основные понятия систем поддержки принятия решений и OLAP, Основные задачи Data Mining. Найти в Internet, например, на сайте http://www.basegroup.ru/ примеры использования Data Mining.	1–5	1
1	1.2	Подготовка к лабораторным работам	Изучить рабочую среду Deductor Academic.	4	1
2	1.3	Подготовка к лекциям	Изучить требования, предъявляемые к базам данных в СППР, концепцию хранилища данных, различные реализации OLAP-систем.	1–5	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
			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.		
2	1.4	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
3	2.1	Подготовка к лекциям	Изучить виды трансформации данных: преобразование временных данных, группировка и слияние, квантование, нормализация.	1–5	1
3	2.2	Подготовка к лабораторным работам	Разработать структуру хранилища данных. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
4	2.3	Подготовка к лекциям	Изучить основные операции предобработки данных: обработка дубликатов, противоречий и аномалий, восстановление пропущенных данных, сокращение размерности исходных данных.	1–5	1
4	2.4	Подготовка к лабораторным работам	Освоить работу с хранилищем данных Deductor Studio. Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.	4	1
		Выполнение индивидуального задания	Анализ задания, подбор литературы.	1–5	2
5	3.1	Подготовка к лекциям	Изучить понятие ассоциативного правила и числовые характеристики правила: поддержка, достоверность, лифт, леввередж.	1–5	1
5	3.2	Подготовка к лабораторным работам	Изучить алгоритмы предварительной обработки данных в Deductor Studio. 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.	1–5	2
6	3.3	Подготовка к лекциям	Изучить алгоритм Apriori.	1–5	1
6	3.4	Подготовка к лабораторным работам	Изучить поиск ассоциативных правил в Deductor Studio. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Выбор методов анализа для конкретного задания.	1–5	2
7	4.1	Подготовка к лекциям	Изучить алгоритм "1 правило" построения правил классификации. Изучить байесовскую классификацию, "наивно"-байесовский подход (Naive-Bayes approach).	1–5	1
7	4.2	Подготовка к лабораторным работам	Изучить логистическую регрессию. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Доработка перечня методов анализа учетом замечаний преподавателя.	1–5	2
8	4.3	Подготовка к лекциям	Изучить выбор атрибута ветвления дерева решений, алгоритмы ID3, C4.5.	1–5	1
8	4.4	Подготовка к ла-	Изучить реализацию алгоритма классификации в Deductor Studio.	4	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
		бораторным работам	Оформление отчета о лабораторной работе. Подготовка к защите лабораторной работы.		
		Выполнение индивидуального задания	Предварительная обработка данных задания.	4	2
9	4.5	Подготовка к лекциям	Изучить алгоритм CART, понятие регрессионного дерева решений, принципы упрощения деревьев решений.	1–5	1
9	4.6	Подготовка к лабораторным работам	Повторить структуру и принципы обучения многослойного перцептрона. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
10	4.7	Подготовка к лекциям	Изучить использование нейронных сетей для решения задач классификации и регрессии.	1–5	1
10	4.8	Подготовка к лабораторным работам	Изучить реализацию нейросетевого прогнозирования в Deductor Studio. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	1–5	2
11	5.1	Подготовка к лекциям	Изучить меры близости, используемые в кластеризации, иерархические алгоритмы и неиерархические алгоритмы кластеризации.	1–5	1
11	5.2	Подготовка к лабораторным работам	Изучить реализацию алгоритма k-means. Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Реализация выбранных алгоритмов Data Mining. Анализ результатов.	4	2
		Подготовка к контрольной работе	Повторить основные алгоритмы Data Mining.	1–5	2
12	5.3	Подготовка к лекциям	Изучить построение сетей и карт Кохонена.	1–5	1
12	5.4	Подготовка к лабораторным работам	Изучить реализацию в Deductor алгоритма кластеризации с помощью карт Кохонена.	4	1
		Выполнение индивидуального задания	Оформление отчета о выполнении индивидуального задания.		2
13	6.1	Подготовка к лекциям	Изучить понятие и модели временного ряда. Изучить корреляционный анализ временных рядов.	1–5	1
13	6.2	Подготовка к лабораторным работам	Изучить анализ временных рядов. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Оформление отчета о выполнении индивидуального задания. Подготовка к защите отчета.	4	2
14	6.3	Подготовка к лекциям	Изучить прогнозирование временного ряда методами среднего и скользящего среднего. Изучить регрессионные модели прогнозирования и моделирование временного ряда случайным процессом.	1–5	1

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
14	6.4	Подготовка к лабораторным работам	Подготовка к защите лабораторной работы. Оформление отчета о лабораторной работе.	4	1
		Выполнение индивидуального задания	Защита отчета.	1–5	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 материал. 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-5, 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.

Демонстрационный вариант контрольной работы

Используя файл тестовых данных `loans.txt` фирмы BaseGroup Labs, содержащий информацию о заемщиках банка, с помощью Deductor построить дерево решений, реализующее кредитный скоринг, то есть относящих потенциального заемщика к одному из двух классов — "плохой", которому кредит не выдается, и "хороший", которому кредит выдается.

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

7.1. Основная литература

1.	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
2.	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)
3.	Чубукова И. А. Data Mining. — НОУ Интуит, 2016. — 470 с. (ЭБС book.ru http://www.book.ru/book/917500)

7.2. Дополнительная литература

4.	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
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 ресурсы на русском языке.

№ п/п	Адрес сайта	Описание материала, содержащегося на сайте
1.	www.olap.ru	Российский сайт, посвященный Business Intelligence и OLAP-технологии. Содержит новости и статьи по указанной тематике.
2.	www.basegroup.ru	Сайт фирмы BaseGroup Labs. Содержит много материалов по методологии Data Mining и системе Deductor. Можно скачать свободно распространяемую версию Deductor Academic.
3.	http://www.intuit.ru/department/database/datamining/	Чубукова И. А. Курс: Data Mining
4.	http://www.intuit.ru/department/database/datawarehouse/	Перминов Г. И. Видеокурс: Хранилища данных
5.	http://www.intuit.ru/studies/courses/3498/740/info	Видеокурс: Алгоритмы интеллектуальной обработки больших объемов данных.
6.	http://www.machinelearning.ru	MachineLearning.ru — профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
	https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie	Интерактивный курс Введение в машинное обучение Школы анализа данных Яндекс:
	https://yandexdataschool.ru/edu-process/courses/machine-learning#item-1	Видеолекции "Машинное обучение" Школы анализа данных Яндекс.
	https://ru.coursera.org/specializations/machine-learning	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 и индивидуальные задания. 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 № 8^а от "16" сентября 20 15 года

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

Программа одобрена методической комиссией факультета ВТ


протокол № 5^а от «17» сентября 20 15 г.

Председатель методической комиссии факультета ВТ


(подпись)

Качнов Н.Н.
(Ф.И.О.)

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

Учеб- ный год	Решение кафедры (№ протокола, дата, подпись зав. кафед- рой)	Внесенные измене- ния	Номера листов		
			заменен- ных	новых	аннулиро- ванных
2017/ 2018	Переутвердить №1 от 31.08.2017 	—	—	—	—