

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет автоматики и вычислительной техники

“УТВЕРЖДАЮ”

Декан АВТФ

профессор, д.т.н. Гужов
Владимир Иванович

“ ___ ” _____ г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Интеллектуальный анализ данных

ООП: специальность 230101.65 Вычислительные машины, комплексы, системы и сети

Шифр по учебному плану: ДС.Ф.6

Факультет: автоматики и вычислительной техники очная форма обучения

Курс: 5, семестр: 9

Лекции: 30

Практические работы: - Лабораторные работы: 30

Курсовой проект: - Курсовая работа: - РГЗ: 9

Самостоятельная работа: 10

Экзамен: - Зачет: 9

Всего: 70

Новосибирск

2011

Рабочая программа составлена на основании Государственного образовательного стандарта высшего профессионального образования по направлению (специальности): 654600 Информатика и вычислительная техника.(№ 224 тех/дс от 27.03.2000)

ДС.Ф.6, дисциплины федерального компонента

Рабочая программа обсуждена на заседании кафедры Вычислительной техники протокол № 6 от 31.08.2011

Программу разработал

доцент, к.т.н.

Зыбарев Владимир Михайлович

Заведующий кафедрой

профессор, д.т.н.

Губарев Василий Васильевич

Ответственный за основную образовательную программу

профессор, д.т.н.

Губарев Василий Васильевич

1. Внешние требования

Таблица 1.1

Шифр дисциплины	Содержание учебной дисциплины	Часы
ДС.Ф.6	<p>Требования к профессиональной подготовленности дипломированного специалиста. (п 7.1)</p> <p>Инженер по специальности "Вычислительные машины, комплексы, системы и сети"</p> <p>должен знать:</p> <ul style="list-style-type: none"> - методы, технологии и инструментальные средства, применяемые на всех этапах разработки аппаратно-программных комплексов; - формальные модели, применяемые при анализе, разработке и испытаниях аппаратно-программных комплексов; <p>должен владеть:</p> <ul style="list-style-type: none"> - методами и средствами теоретического и экспериментального исследования, ориентированными на создание перспективных средств ВТ; - методами разработки и анализа алгоритмов, моделей, архитектур и структур аппаратно-программных комплексов; - методами и средствами анализа аппаратно-программных комплексов, методами метрологии и обеспечения качества их функционирования; - методами и средствами анализа, описания и проектирования человеко-машинного взаимодействия, инструментальными средствами разработки пользовательского интерфейса. <p>Концептуальная записка по направлению подготовки / специальности 230101.65 "Вычислительные машины, комплексы, системы и сети"</p> <p>Основы анализа и интерпретации данных (знаний). Методы прикладного и статистического анализа данных. Анализ многомерных данных. Основные задачи , методы и средства интеллектуального анализа данных. Компьютерные технологии и системы анализа данных.</p>	70

2. Особенности (принципы) построения дисциплины

Особенности (принципы) построения дисциплины

Особенность (принцип)	Содержание
Основания для введения дисциплины в учебный план по направлению или специальности	<p>Государственный образовательный стандарт высшего профессионального образования, направление подготовки дипломированного специалиста 654600 - "Информатика и вычислительная техника" от 27.03.2000, регистрационный номер 224 тех/дс;</p> <p>Цикл СД - Специальные дисциплины</p> <p>Решение Ученого совета факультета автоматизи и вычислительной техники протокол №4 от 18.04.2007г.</p>
Адресат курса	<p>Подготовка дипломированных специалистов по специальности</p> <p>230101 - "Вычислительные машины, комплексы, системы и сети"</p>
Основная цель (цели) дисциплины	<p>Курс входит в число дисциплин цикла СД - Специальные дисциплины.</p> <p>Основной целью курса является изучение некоторых современных информационных технологий, предназначенных для интеллектуального анализа данных, направленных на формирование целостного представления об анализе и интерпретации экспериментальных и статистических данных, как о процессе поиска, так и применения скрытых в них закономерностей для достижения поставленных целей.</p>
Ядро дисциплины	<p>Ядро курса составляют средства и методы прикладного статистического анализа, а так же специальные алгоритмы, обеспечивающие извлечение знаний из данных.</p> <p>В курсе закрепляется и развивается представление о системном подходе к сбору, обработке и использованию результатов анализа экспериментальных данных для создания математических моделей адекватных окружающему миру и решения прикладных задач.</p>
Связи с другими учебными дисциплинами основной образовательной программы	<p>Для успешного изучения курса студенту необходимо знать основы: информатики, алгебры, математического анализа, дискретной математики, теории вероятностей, математической статистики, случайных процессов, баз данных, систем искусственного интеллекта, методов оптимизации и принятия решения;</p>
Требования к первоначальному уровню подготовки обучающихся	<p>Умение работать с операционными системами (например Microsoft Windows), с программными продуктами Microsoft Office или другими офисными приложениями и базами данных. Владение приемами и средствами вероятностного и статистического анализа.</p>
Особенности организации учебного процесса по дисциплине	<p>Методическую основу учебного курса составляют лекции, лабораторные работы, РГЗ, аудиторные и самостоятельные занятия. Лекционные занятия раскрывают основные проблемные вопросы по каждому разделу.</p> <p>Лабораторные работы и РГЗ предназначены для приобретения практических навыков в работе по освоению и применению различных методов анализа данных в прикладных</p>

	<p>информационных системах и в экономике. Выполняя практическую часть курса, студенты применяют теоретические положения анализа данных, приобретают практические навыки использования методов интеллектуального анализа для обработки экспериментальных данных. Часть типовых заданий может меняться с учетом индивидуальных склонностей студента и содержания его будущей выпускной работы. Приветствуется использование в качестве рабочего материала реальных данных, получаемых самим студентом при проведении своих исследований. В качестве РГЗ может также быть выполнена эмпирическая (аналитическая) часть диплома, если в ней используются обсуждаемые на данном курсе методы и средства анализа и интерпретации данных.</p> <p>Для проведения лабораторных работ используются сборники типовых заданий, методические указания и рекомендации по применению средств специализированных программных систем анализа данных. Оценка знаний и умений студентов проводится на основе результатов выполненного лабораторного практикума и РГЗ, осуществляемая в форме индивидуальной защиты отчетов о проделанной работе и результатов анализа.</p> <p>Самостоятельная работа студентов проводится вне рамок учебного расписания и предназначена для закрепления лекционного материала, знакомства с рекомендованной литературой, подготовки к лабораторным работам и выполнения РГЗ. Формами контроля самостоятельной работы являются факультативные семинары, консультации, индивидуальные задания и зачет.</p>
--	---

3. Цели учебной дисциплины

Таблица 3.1

После изучения дисциплины студент будет

иметь представление	
1	О роле и целях применения компьютеризированного анализа и интерпретации данных в познании окружающего мира и решения практических задач.
2	О прикладных методах анализа данных и добычи знаний.
3	Основные направления и тенденции их развития,
4	О возможностях и перспективах применения методов интеллектуального анализа данных в исследовании объектов и процессов профессиональной деятельности в области информатики и ВТ.
5	Об организации и применении современных информационных технологий и программных систем для обработки экспериментальных и статистических данных.
знать	
6	Предмет и объекты изучения, структуру данного курса, сущность используемых в нем основополагающие методов исследований, место его в учебном плане для подготовки дипломированных специалистов по данному направлению.
7	Базовые определения и понятия, основные этапы анализа и интерпретации данных, общую структуру и функционирование интегрированных информационно-аналитических систем (систем поддержки принятия решений).
8	Основы организации и технологии обработки экспериментальных и статистических данных на ЭВМ.
9	Содержательную и математическую постановку основных задач анализа данных, способы их решения.
10	Сущность и свойства основных алгоритмов интеллектуального анализа
уметь	
11	Ориентироваться в области теории вероятностей и математической статистики, вычислительной математики и компьютерной графики, пользоваться специальной литературой в изучаемой предметной области;
12	Вести дискуссию в предметных областях, сопряженных с анализом данных, в том числе обосновывать выбор средств и методов для решения конкретных задач учебного назначения;
13	Сводить словесные постановки задач к типовым математическим и относить их к соответствующим разделам математики, связывать их с известными средствами обработки экспериментальных данных;
14	Готовить данные к анализу и проводить их визуализацию
иметь опыт (владеть)	
15	Создавать и пользоваться типовыми математическими моделями для исследования случайных явлений и процессов по результатам наблюдений при решении задач: сравнения, выявления и восстановления

	закономерностей, классификации, прогнозирования;
16	Ставить задачи, выбирать средства и методы обработки экспериментальных данных для эффективного применения;
17	Представлять результаты анализа в удобном для восприятия виде, интерпретировать их в соответствии с поставленной задачей;
18	Осуществлять самооценку и самоконтроль при решении задач обработки статистических данных.

4. Содержание и структура учебной дисциплины

Лекционные занятия

Таблица 4.1

(Модуль), дидактическая единица, тема	Часы	Ссылки на цели
Семестр: 9		
Модуль: Введение		
Дидактическая единица: Предмет и объект изучения, характеристика целей и задач ИАД		
Введение	2	1, 6
Модуль: Основные понятия дисциплины		
Дидактическая единица: Способы задания данных, извлечения и представления знаний		
Основы анализа данных и способы извлечения знаний	2	2, 6, 7, 8
Модуль: Анализ многомерных данных и кластерный анализ		
Дидактическая единица: Анализ и визуализация многомерных данных при факторном и кластерном анализе данных		
Методы и задачи факторного и кластерного анализа	4	13, 17, 3, 6, 9
Модуль: Классификация данных на основе дискриминантных функций		
Дидактическая единица: Способы задания, создания и применения дискриминантных функций в решении задач классификации		
Решение задач дискриминантного анализа	4	11, 15, 16, 17, 2, 6, 7, 8
Модуль: Обнаружение логических закономерностей и деревья решений.		
Дидактическая единица: Методы анализа разнотипных данных и применение деревьев решений для изучения закономерностей		
Применения деревьев решений для анализа данных	4	10, 15, 17, 3, 9
Модуль: Основные этапы и задачи анализа данных при выявлении и восстановлении зависимости.		
Дидактическая единица: Методы и способы решения задач обнаружения и восстановления зависимостей по эмпирическим данным		
Методы и способы выявления и восстановления зависимости на основе анализа данных.	4	11, 12, 13, 15, 2, 5, 8, 9

Модуль: Анализ и прогнозирование временных рядов.		
Дидактическая единица: Применение математических моделей для анализа временных рядов и прогнозирования		
Методы и способы анализа и прогнозирования временных рядов.	4	13, 15, 16, 2, 8, 9
Модуль: Извлечение знаний из данных (Системы DATA MINING)		
Дидактическая единица: Системы и технологии извлечения знаний из данных		
Системы DATA MINING в задачах анализа и интерпретации данных.	4	10, 16, 18, 3, 4, 5
Модуль: Современные пакеты прикладных программ для решения задач обработки экспериментальных данных		
Дидактическая единица: Системы и технологии извлечения знаний из данных		
Организация и применение пакетов прикладных программ для решения задач обработки экспериментальных данных	2	12, 16, 17, 18, 2, 5, 8
Модуль: Заключение		
Дидактическая единица: Состояние и перспективы развития и применения ИАД		
Обзор состояния и перспектив ИАД	2	1, 3, 4

Лабораторная работа

Таблица 4.2

(Модуль), дидактическая единица, тема	Учебная деятельность	Часы	Ссылки на цели
Семестр: 9			
Модуль: Основные этапы и задачи анализа данных при выявлении и восстановлении зависимости.			
Дидактическая единица: Методы и способы решения задач обнаружения и восстановления зависимостей по эмпирическим данным			
Сравнительный анализ объектов и факторных эффектов	Познакомиться и приобрести опыт решения задачи сравнения объектов и факторных эффектов по результатам наблюдений методами математической статистики и средствами визуализации данных	4	1, 11, 12, 13, 14, 15, 16, 17, 2, 7

Выявления и восстановление функциональных зависимостей средствами регрессионного анализа.	Создание по экспериментальным данным математических моделей, описывающих поведение некоторой характеристики в зависимости от изменения множества признаков, анализ значимости факторов и связей между ними средствами корреляционно-регрессионного анализа.	4	1, 11, 13, 15, 16, 17, 18, 8
Модуль: Анализ многомерных данных и кластерный анализ			
Дидактическая единица: Анализ и визуализация многомерных данных при факторном и кластерном анализе данных			
Кластерный анализ	Исследование совокупности объектов (данных), представляемой в виде матрицы "Объект-Свойства" методом разбиения этой совокупности на однородные группы (кластеры) схожих по свойствам объектов на минимальном наборе дискриминантных характеристик.	4	12, 13, 14, 17, 5
Модуль: Классификация данных на основе дискриминантных функций			
Дидактическая единица: Способы задания, создания и применения дискриминантных функций в решении задач классификации			
Дискриминантный анализ	Определение принадлежности объекта к одной из групп, которые заданы обучающими выборками в виде множества образцовых примеров классификации объектов.	4	13, 14, 15, 16, 17, 18

Модуль: Обнаружение логических закономерностей и дерева решений.			
Дидактическая единица: Методы анализа разнотипных данных и применение деревьев решений для изучения закономерностей			
Применение деревьев решения для классификации (сегментации) данных.	Выделение сегментов и обнаружения скрытых тенденций в многомерных данных на основе логических правил и древовидных диаграмм.	4	10, 14, 17, 18, 2, 4
Модуль: Анализ и прогнозирование временных рядов.			
Дидактическая единица: Применение математических моделей для анализа временных рядов и прогнозирования			
Анализ временных рядов и прогнозирование	Знакомство с процедурами сглаживания и приобретение опыта разложения временного ряда на составляющие для создания математических моделей и прогнозирования.	4	13, 14, 15, 16, 17, 18, 2, 8, 9
Модуль: Извлечение знаний из данных (Системы DATA MINING)			
Дидактическая единица: Системы и технологии извлечения знаний из данных			
Применение нейронных сетей для решения основных задач ИАД.	Решение задач восстановления функциональной зависимости, классификации и прогнозирования на основе нейронных сетей и сравнение полученных результатов с полученными другими методами при выполнении предыдущих лабораторных работ.	6	14, 17, 18, 2, 4

5. Самостоятельная работа студентов

Семестр- 9, Подготовка к зачету

Подготовка к зачету осуществляется путем изучения теоретического материала на основе прослушенного курса лекций и рекомендованных литературных источников. Важной составляющей подготовки к зачету является систематизация и анализ результатов выполнения индивидуальных практических заданий по анализу данных и лабораторному практикуму.

Объем самостоятельной работы студента 10 часов

Семестр- 9, РГЗ

Цель задания: приобретение опыта применения новых информационных технологий и программных систем, предназначенных для проведения интеллектуального анализа данных; развитие умения устанавливать эмпирические факты, формировать рабочие гипотезы, выбирать средства обработки экспериментальных данных, формы представления и интерпретации результатов анализа; формирование целостного представления об анализе и интерпретации экспериментальных и статистических данных, как о процессе поиска, так и применения скрытых в них закономерностей для достижения поставленных целей при решении практических задач.

Индивидуальные задания могут быть:

типовыми;

исследовательскими;

связанными с подготовкой реферата по одному из разделов курса или не вошедших в него, но актуальных и значимых для текущей практической деятельности студента и дипломированного специалиста;

направлены на изучение и освоение современных средств и систем анализа данных (информационных технологий) в сфере профессиональной деятельности специалиста;

являться в будущем собственной разработкой и экспериментальной частью выпускной работы (дипломного проекта).

Тематика задания: Тема индивидуального задания выбирается студентом как из предложенного преподавателем списка прикладных задач информатики и применения средств вычислительной техники в различных предметных областях деятельности дипломированных специалистов, в частности в экономике, так и в порядке личной инициативы, при условии что она связана с использованием изучаемых в данном курсе методов и средств ИАД. Часть типовых заданий может меняться с учетом индивидуальных склонностей студента и содержания его будущей выпускной работы. Приветствуется использование в качестве рабочего материала реальных данных, получаемых самим студентом при проведении своих исследований. В качестве РГЗ может также быть выполнена эмпирическая (аналитическая) часть диплома. Содержание индивидуального задания определяется по результатам собеседования преподавателя со студентом не позже месяца с начала учебных занятий.

Содержание работы:

- 1) Знакомство с литературой и примерами выполненного анализа данных по выбранной тематике.
- 2) Изучение заданных или сбор исходных данных (Постановка прикладной задачи).
- 3) Определение целей анализа данных для решения прикладной задачи.
- 4) Подготовка, предварительная обработка и качественный анализ данных.
- 5) Определение рабочих моделей данных и применение методов и средств ИАД для формирования и аргументации результатов анализа данных .
- 6) Оценка достоверности и применение (интерпретация) результатов анализа для решения прикладной задачи.

7) Выбор формы и написание отчета о проделанной работе, о результатах анализа и их интерпретации.

Требования к оформлению пояснительной записки

Пояснительная записка может быть представлена на листах формата А4 объемом не более 25 страниц. Содержание пояснительной записки должно состоять из введения; индивидуального задания; краткого описания применяемых методик, средств ИАД и технологии проведения анализа данных; расчетно-аналитическая часть с комментариями (результаты вычислений, графики, таблицы, анализ и интерпретация результатов); заключение (краткое изложение результатов работы, выводы по результатам анализа модели, возможности ее использования при исследовании конкретного объекта, рекомендации по применению средств ИАД и технологии анализа данных); ответы на контрольные вопросы, заданные преподавателем; приложения (алгоритмы, программы, сценарии анализа, инструкции по применению средств ИАД и прочее).

Подробнее о РГЗ и примеры выполнения содержатся в методических указаниях, выдаваемых студентам на электронном носителе.

Объем самостоятельной работы студента 20 часов.

Семестр- 9, Подготовка к занятиям

Самостоятельная работа студентов проводится вне рамок учебного расписания и предназначена для закрепления лекционного материала, поиск и изучение рекомендованной литературой, подготовки к лабораторным работам и выполнения РГЗ. При подготовке как к лекциям, так и лабораторным работам основой самостоятельной работы является знакомство с типовыми решениями наиболее распространенных задач анализа данных и экспериментирование с образцовыми исходными данными, средствами и методами ИАД. Формами контроля самостоятельной работы являются факультативные семинары, консультации, график выполнения индивидуальных заданий.

Объем самостоятельной работы студента 40 часов.

6. Правила аттестации студентов по учебной дисциплине

Студенты по учебной дисциплине аттестуются в виде зачета по результатам контроля знаний в форме индивидуальных защит отчетов по выполненному лабораторному практикуму, РГЗ и ответов на контрольные вопросы по материалам лекционного курса. В порядке исключения допускается аттестация студента по результатам выполненного индивидуального задания, определенного преподавателем или выбранного студентом по личной инициативе, при условии его публичной защиты.

Для аттестации студентов по дисциплине используется балльно-рейтинговая система. Рейтинг студента по дисциплине определяется как сумма баллов за работу в семестре (текущий рейтинг) и баллов, полученных в результате итоговой аттестации (зачет).

Максимальная сумма баллов за семестр составляет 100 баллов (текущий рейтинг - 60 баллов, итоговая аттестация - 40 баллов).

Правила текущей аттестации:

1. В течение шестого семестра необходимо представить и защитить 8 лабораторных работ, расчетно-графическую работу в сроки, установленные учебным графиком (см. таблицу).

2. К защите допускаются студенты, выполнившие лабораторные работы, РГР в полном объеме (все задания согласно варианту) и оформившие отчет по работе в соответствии с требованиями.

3. На защите предлагается два теоретических вопроса и один практический вопрос (по ходу выполнения работы).

4. Максимальное количество баллов (6 или 11-12 в зависимости от вида работы) выставляется, если студент полностью ответил на все вопросы, без серьезных замечаний и недочетов.

5. Количество баллов 4-5 или 8-10 (в зависимости от вида работы) выставляется, если студент полностью ответил на два вопроса из трех, причем один из вопросов - практический.

6. Минимальное количество баллов 3 или 6-7 (в зависимости от вида работы) выставляется, если студент ответил на два вопроса из трех частично, с серьезными замечаниями, недочетами.

7. Пересдача лабораторной работы, РГР назначается, если студент не ориентируется в учебном материале, не может объяснить ход и результаты выполнения работы. В случае пересдачи работы происходит потеря баллов (максимальное количество баллов составляет 5 или 10 в зависимости от вида работы).

8. В случае представления и защиты работ с опозданием от учебного графика происходит потеря баллов (опоздание на 1 неделю - потеря 1 или 2 баллов в зависимости от вида работы, опоздание на 2 недели - потеря 2 или 4 баллов, 3 недели и более - потеря 50% баллов от максимально возможного).

Правила итоговой аттестации:

1. К зачету допускаются студенты, сдавшие лабораторные работы, РГР, и набравшие не менее 50% (30 баллов) по результатам текущего рейтинга.

2. Зачет проводится в устном виде, предлагается три теоретических вопроса (образцы вопросов приведены в п.12).

3. Максимальное количество 36-40 баллов выставляется, если все задания выполнены полностью, без серьезных замечаний.

4. Количество баллов 30-35 выставляется, если успешно выполнены два задания из трех, причем один из отвеченных вопросов - теоретический.

5. Минимальное количество баллов 20-29 выставляется, если выполнены два задания из трех, но с серьезными ошибками, замечаниями, недочетами.

6. Возможно получить "автомат" (зачтено) по дисциплине без сдачи зачета, если студент в течение семестра набирает свыше 90 баллов по текущему рейтингу.

7. Список литературы

7.1 Основная литература

В печатном виде

1. Загоруйко Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. - Новосибирск, 1999. - 269 с. : ил. - Библиогр.: с. 247-260. - Предм. указ.: с. 261-263.
2. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл.
3. Чубукова И. А. Data Mining : учебное пособие / И. А. Чубукова. - М., 2006. - 382 с. : ил.
4. Цильковский И. А. Методы анализа знаний и данных : конспект лекций / И. А. Цильковский, В. М. Волкова ; Новосиб. гос. техн. ун-т. - Новосибирск, 2010. - 66, [2] с. : ил.
5. Боровиков В. П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учебное пособие для вузов по специальности "Прикладная математика" / В. П. Боровиков, Г. И. Ивченко. - М., 2006. - 367, [1] с. : ил. - Рекомендовано МО.

В электронном виде

1. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл. - Режим доступа: <http://www.ciu.nstu.ru/fulltext/textbooks/2008/avdeenko.pdf>. - Инновационная образовательная программа НГТУ «Высокие технологии».
2. Цильковский И. А. Методы анализа знаний и данных : конспект лекций / И. А. Цильковский, В. М. Волкова ; Новосиб. гос. техн. ун-т. - Новосибирск, 2010. - 66, [2] с. : ил.

7.2 Дополнительная литература

В печатном виде

1. Елисеева И. И. Практикум по общей теории статистики : [практикум для вузов по специальности "Статистика" и другим экономическим специальностям] / И. И. Елисеева, Н. А. Флуд, М. М. Юзбашев ; под ред. И. И. Елисеевой. - М., 2008. - 509, [1] с. - Рекомендовано УМО.
2. Ларсон Б. Разработка бизнес-аналитики в Microsoft SQL Server 2005 / Б. Ларсон ; [пер. с англ. А. Конев, В. Щербинин]. - М. [и др.], 2008. - 683 с. : ил.
3. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных : [наиболее полное руководство в подлиннике / Александр Бергер и др.] ; под общ. ред. Александра Бергера, Ирины Горбач. - СПб., 2007. - 905 с. : ил., портр.
4. Айвазян С. А. Прикладная статистика. Основы эконометрики. Т. 1. Теория вероятностей и прикладная статистика/ С. А. Айвазян, В. С. Мхитарян : Учебник для экон. спец. вузов: В 2 т.. - М., 2001. - 656 с.
5. Айвазян С. А. Прикладная статистика. Основы эконометрики. Т. 2. Основы эконометрики/ С. А. Айвазян : Учебник для экон. спец. вузов: В 2 т.. - М., 2001. - 432 с.
6. Боровиков В. П. STATISTICA: Статистический анализ и обработка данных в среде Windows / В. П. Боровиков, И. П. Боровиков. - М., 1997. - 608 с.
7. Дуброва Т. А. Статистические методы прогнозирования : учебное пособие для вузов по специальности 061700 "Статистика" и др. экон. специальностям / Т. А. Дуброва. - М., 2003. - 206 с. : ил. - Рекомендовано УМО.
8. Практикум по эконометрике : Учеб. пособие для экон. вузов / [И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др.]; Под ред. И. И. Елисеевой. - М., 2001. - 191 с.

9. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров; Под ред. В. Э. Фигурнова. - М., 2003. - 544 с. : ил.
10. Базы данных : интеллектуальная обработка информации / В. В. Корнеев и др.; Рос. Ассоц. изд. компьютер. лит. - М., 2000. - 351 с. : ил., табл.
11. Лбов Г. С. Анализ данных и знаний : учебное пособие [для 3-5 курсов ФПМИ (специальность 5102) дневного отделения] / Г. С. Лбов ; Новосиб. гос. техн. ун-т. - Новосибирск, 2001. - 86 с. : ил.
12. Макленнен Д. Microsoft SQL Server 2008: Data Mining : интеллектуальный анализ данных : [наиболее полное руководство] / Джемс Макленнен, Чжаохуэй Танг, Богдан Криват ; [пер. с англ. А. Лашкевича]. - СПб., 2009. - XX, 700 с. : ил.
13. Плис А. И. Практикум по прикладной статистике в среде SPSS. Ч. 1 : учебное пособие для вузов / А. И. Плис, Н. А. Сливина. - М., 2004. - 287 с. : ил. + 1 CD-ROM. - Рекомендовано МО.
14. Дубнов П. Ю. Обработка статистической информации с помощью SPSS / Дубнов П. Ю. - М., 2004. - 221 с. : ил. - На обл.: получение и оценка статистических параметров и гипотез ; обработка статистических выборок ; анализ временных рядов..
15. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям : учебное пособие / Н. Б. Паклин, В. И. Орешков. - СПб. [и др.], 2010. - 701 с. : ил. + 1 CD-ROM.

8. Методическое и программное обеспечение

8.1 Методическое обеспечение

В печатном виде

1. Решение задач интеллектуального анализа данных на основе вариативного моделирования средствами пакета Statistica : методические указания к лабораторным работам по дисциплине "Интеллектуальный анализ данных" для 5 курса АВТФ / Новосиб. гос. техн. ун-т ; [сост. Альсова О. К.]. - Новосибирск, 2005. - 75 с. : табл.
2. Обнаружение и анализ закономерностей в эмпирических данных : методическое пособие для 5 курса ФПМИ / Новосиб. гос. техн. ун-т ; [сост.: Г. С. Лбов, В. М. Неделько]. - Новосибирск, 2008. - 19, [1] с. : табл.

В электронном виде

1. Обнаружение и анализ закономерностей в эмпирических данных : методическое пособие для 5 курса ФПМИ / Новосиб. гос. техн. ун-т ; [сост.: Г. С. Лбов, В. М. Неделько]. - Новосибирск, 2008. - 19, [1] с. : табл. - Режим доступа: <http://www.library.nstu.ru/fulltext/metodics/2008/3585.rar>

8.2 Программное обеспечение

1. Statsoft, Statsoft STATISTICA , универсальное средство анализа данных

9. Контролирующие материалы для аттестации студентов по дисциплине

Примерный перечень контрольных вопросов:

1. Основные цели и задачи ИАД.
2. Средства и методы ИАД.
3. Назначение, основные этапы развития и виды систем ИАД.
4. Применение систем ИАД для информационно-аналитического обеспечения принятия решений (на основе выбранного примера).
5. Основные формы представления и способы объяснения (интерпретации) результатов ИАД.
6. Основные этапы анализа и интерпретации данных
7. Стратегии ввода, представления и организация исходных данных в информационных системах с ИАД.
8. Измерительные шкалы, представление переменных, ввод и редактирование данных.
9. Трансформация данных и файлов в методах ИАД (на выбранном методе, примере решения задачи).
10. Основные способы загрузки данных в системы ИАД из различных источников, в том числе на основе ODBC.
11. Объединение и агрегирование (файлов) данных в информационно-аналитических системах с применением средств ИАД
12. Базовые разведочные методы анализа данных.
13. Способы визуализации информации в системах ИАД
14. Системы для визуализации многомерных данных
15. Графические способы представления результатов ИАД.
16. Табличные способы представления результатов ИАД.
17. Применение методов и средств картографии для анализа данных и интерпретации его результатов.
18. Построение и интерпретации карт визуализации зависимостей (на выбранном примере, в том числе - анализа категориальных данных).
19. Построение, применение и интерпретации карт категориальных данных.
20. Трансформируемые таблицы и интерактивные диаграммы (на примере "сводных таблиц" MS Office).
21. Подготовка отчетов и экспорт результатов анализа из систем ИАД.
22. Классические методы многомерного статистического анализа.
23. Роль и место методов классического статистического анализа для решения основных задач ИАД.
24. Возможности программного обеспечения и практическое применение инструментов классического статистического анализа данных для решения задач ИАД (на примере выбранной системы или пакета прикладных программ).
25. Задачи выявления и восстановления зависимостей в ИАД.
26. Простая линейная и многомерная регрессия
27. Нелинейная регрессия
28. Бинарная и множественная логистическая регрессия.
29. Задачи и методы классификации в ИАД.
30. Применение методов классификации и кластеризации для сегментации данных.
31. Дискриминационный анализ как способ классификации явлений и объектов.
32. Принципы и общая характеристика методов кластерного анализа.
33. Применение кластерного анализа для сокращения количества переменных при моделировании (на выбранном примере).
34. Использование деревьев решений в задачах классификации (на выбранном примере).
35. Общая характеристика и методы решения задач прогнозирования
36. Построение прогнозов на основе различных моделей.
37. Оценивание качества прогноза и адекватности модели прогнозирования.

38. Анализ и прогнозирование временных рядов: цели, задачи, методы (временной и частотный подходы к анализу временных рядов).
39. Компоненты и модели временных рядов (тренда, сезонности, цикла, случайной компоненты).
40. Обзор моделей экспоненциального сглаживания и ARIMA.
41. Использование моделей временных рядов для анализа данных и прогнозирования (на выбранном примере).
42. Способы декомпозиции временных рядов: выявления тренда, сезонной, циклической и случайных составляющих (на выбранном примере).
43. Введение в нейронные сети: основные принципы организации, построения и применения.
44. Решение задач классификации нейронными сетями (на выбранном примере).
45. Применение нейронных сетей к задачам регрессионного анализа (на выбранном примере).
46. Применение нейронных сетей к задачам прогнозирования временных рядов (на выбранном примере).
47. Деревья решений, принципы их построения, практического применения и интерпретации результатов (характеристика деревьев).
48. Анализ с помощью дерева решений: обычные задачи анализа с помощью дерева решений, области приложения анализа с помощью дерева решений.
49. Применение дерева решений для классификации (на выбранном примере).
50. Применение дерева решений для прогнозирования (на выбранном примере).
51. Сравнительный анализ средств (систем) анализа и прогнозирования на основе деревьев решений (на примере не менее трех систем).
52. Автоматизация выполнения повторяющихся задач в системах ИАД.
53. Объединение и агрегирование (файлов) данных в информационно-аналитических системах с применением средств ИАД.
54. Создание отчетов и экспорт результатов из систем ИАД в различные приложения (на примере выбранной системы).
55. Влияние размерности и объемов исходных данных на выбор и применение средств ИАД.
56. Методы анализа разнотипных экспериментальных данных.
57. Применение генетических алгоритмов для анализа данных (на примере пакета SNN).
58. Организация и реализация оптимального выбора структуры и параметров моделей в ИАД (на выбранном примере).