

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе

Толстик

Регистрационный № УД- 1187 / уч.



ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальностей:**

1-98 01 01 Компьютерная безопасность (по направлениям)
направления специальности

1-98 01 01-02 Компьютерная безопасность (Радиофизические методы
и программно-технические средства);

1-31 04 02 Радиофизика;

1-31 04 03 Физическая электроника

Учебная программа составлена на основе ОСВО 1-98 01 01-2013, ОСВО 1-31 04 02-2013, ОСВО 1-31 04 03-2013 и учебных планов № Р 98-139/уч. 2013 г., № G 31-164/уч. 2013 г., № G 31-165/уч. 2013 г.

СОСТАВИТЕЛИ:

Николай Николаевич Яцков, доцент кафедры системного анализа и компьютерного моделирования, кандидат физико-математических наук, доцент;

Станислав Викентьевич Гилевский, доцент кафедры системного анализа и компьютерного моделирования, кандидат технических наук, доцент.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой системного анализа и компьютерного моделирования Белорусского государственного университета (протокол № 13 от 22.06.2015 г.);

Научно-методическим советом Белорусского государственного университета (протокол № 6 от 29.06.2015 г.)

A small, square image containing a handwritten signature in black ink on a light green background. The signature is stylized and appears to be the initials 'ВМ'.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная программа «Интеллектуальный анализ данных» (ИАД) разработана для студентов специальностей 1-98 01 01 Компьютерная безопасность (по направлениям) направление специальности 1-98 01 01-02 Компьютерная безопасность (Радиофизические методы и программно-технические средства), 1-31 04 02 Радиофизика и 1-31 04 03 Физическая электроника в соответствии с требованиями ОСВО по специальности 1-98 01 01 «Компьютерная безопасность (по направлениям)» ОСВО 1-98 01 01-2013 и учебного плана БГУ по специальности 1-98 01 01 «Компьютерная безопасность (по направлениям)» направление специальности 1-98 01 01-02 «Компьютерная безопасность (Радиофизические методы и программно-технические средства)», ОСВО по специальности 1-31 04 02 «Радиофизика» ОСВО 1-31 04 02- 2013 и учебного плана БГУ по специальности 1-31 04 02 «Радиофизика», ОСВО по специальности 1-31 04 03 «Физическая электроника» ОСВО 1-31 04 03-2013 и учебного плана БГУ по специальности 1-31 04 03 «Физическая электроника».

Целью учебной дисциплины является изучение теоретических основ технологии ИАД, включая методы статистического анализа, классификации, нейронные сети, кластерного анализа, поиска ассоциативных правил, стохастического поиска, визуализации многомерных данных.

Задачи дисциплины – ознакомление с основными понятиями и типами закономерностей, выявляемыми с помощью ИАД, изучение методов ИАД, обучение использованию программных средств на основе технологии интеллектуального анализа данных для решения практических задач.

Учебная программа составлена с учетом межпредметных связей и программ по учебным дисциплинам «Математический анализ», «Программирование», «Теория вероятностей и математическая статистика».

В результате изучения дисциплины студент должен:

знать:

- базовые понятия и принципы ИАД,
- основные алгоритмы методов и моделей ИАД и подходы к их созданию,
- задачи анализа многомерных массивов данных, этапы процесса ИАД, сферы применения и стандарты ИАД;

уметь:

- производить расчеты с применением аппарата ИАД,
- применять методы ИАД в среде математического программирования Matlab для решения практических задач управления и обработки информации,
- творчески и эффективно использовать полученные знания в профессиональной деятельности.

В соответствии с учебными планами объем дисциплины составляет 140 учебных часов, в том числе 68 аудиторных часов, из них лекции – 34, лабораторные работы – 34.

Дисциплина «Интеллектуальный анализ данных» изучается студентами дневной формы получения высшего образования на 3-ем курсе в 6 семестре.

Форма текущей аттестации по учебной дисциплине – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Введение в интеллектуальный анализ данных. Основные понятия интеллектуального анализа данных. Интеллектуальный анализ данных (ИАД). Методы и задачи дисциплины. Постановка задач ИАД. Данные. Понятие данных. Типы шкал наборов данных. Типы наборов данных. Примеры практического применения дисциплины ИАД. Медицина. Биоинформатика. Геология. Банковское дело. Торговля.

Раздел 2. Статистические методы. Методы предварительного анализа данных. Описательная статистика. Характеристики центральной тенденции. Характеристики вариации. Очистка данных. Нормировка и стандартизация данных. Анализ выбросов и аномальных значений. Корреляционный анализ. Коэффициент корреляции Пирсона. Ранговая корреляция. Критерий Спирмена. Множественная и частная корреляции. Регрессионный анализ. Регрессия. Линейный и нелинейный регрессионный анализ. Дисперсионный анализ. Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ. Методы снижения размерности данных. Метод главных компонент. Факторный анализ.

Раздел 3. Методы классификации. Деревья решений. Анализ данных с использованием деревьев решений. Анализ признаков объектов. Критерии и функции качества разбиения. Индекс Джини. Энтропия. Ошибка классификации. Остановка построения дерева. Сокращение дерева. Обработка пропущенных значений. Извлечение правил из деревьев. Алгоритмы построения деревьев решений. Алгоритм ID3. Алгоритм покрытия. Байесовская классификация. Метод k-ближайших соседей. Метод наименьшего расстояния (до центров кластеров).

Раздел 4. Нейронные сети. Сети классификации и кластеризации данных. Нейрон. Нейронная сеть. Архитектура нейронных сетей. Модель нейронной сети – персептрон. Алгоритмы обучения нейронных сетей. Алгоритм Хебба. Адаптивное обучение. Нейронные сети Кохонена. Нейронная сеть слой Кохонена. Самоорганизующиеся карты Кохонена.

Раздел 5. Кластерный анализ. Основы кластерного анализа. Математические характеристики кластера. Критерии качества кластеризации. Выделение значимых кластеров. Иерархические агломеративные и дивизимные методы иерархического кластерного анализа. Неиерархические итерационные методы кластерного анализа. Алгоритм k-средних. Алгоритм Fuzzy k-средних. Графовые методы кластерного анализа.

Раздел 6. Ассоциативные правила. Методы поиска ассоциативных правил. Ассоциативные правила. Алгоритмы поиска ассоциативных правил. Алгоритм Apriori. Алгоритм FP-G.

Раздел 7. Стохастические методы поиска. Стохастические методы поиска. Методы случайного поиска. Простой стохастический поиск. Случайный поиск. Метод Метрополиса. Метод имитации отжига. Эволюционные стратегии случайного поиска. Селекция. Скрещивание. Мутация. Генетические алгоритмы.

Раздел 8. Методы визуализации данных. Визуальное представление данных. Геометрические проекционные методы. Методы визуализации

многомерных данных. Методы на основе построения иконок. Пиксельные методы визуализации. Принципы качественной визуализации.

Раздел 9. Процесс интеллектуального анализа данных. Основные этапы процесса интеллектуального анализа данных. Анализ предметной области. Постановка задачи. Сбор и подготовка данных. Этапы процесса интеллектуального анализа данных, связанные с построением, проверкой, оценкой, выбором и коррекцией моделей. Анализ данных. Интерпретация и прогнозирование.

Раздел 10. Стандарты и рынок инструментов интеллектуального анализа данных. Организационные и человеческие факторы. Стандарты интеллектуального анализа данных. Инструменты интеллектуального анализа данных: R-пакет, SAS Enterprise Miner, STATISTICA Data Miner, Weka. Интеллектуальный анализ данных с использованием пакета Matlab: интерфейс программы, основные элементы m-языка программирования, анализ многомерных данных.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов				Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	
1	2	3	4	5	6	9
1	Введение в интеллектуальный анализ данных (2 ч)	2	-	-		
1.1	Основные понятия дисциплины «Интеллектуальный анализ данных (ИАД)» 1. Введение 2. Методы, модели и задачи ИАД 3. Данные 4. Примеры применения ИАД					
2	Статистические методы (18 ч)	10	-	-	8	
2.1	Методы предварительного анализа данных 1. Описательная статистика 2. Очистка данных 3. Нормировка и стандартизация данных 4. Анализ выбросов и аномальных значений	2				Тест
2.2	Корреляционный анализ 1. Коэффициент корреляции Пирсона 2. Ранговая корреляция 3. Множественная и частная корреляции	2				
2.3	Регрессионный анализ 1. Регрессия 2. Регрессионный анализ	2			Нелинейный метод наименьших квадратов (4 ч).	Отчет по лабораторной работе
2.4	Дисперсионный анализ 1. Дисперсионный анализ 2. Однофакторный дисперсионный анализ 3. Двухфакторный дисперсионный анализ	2				Тест

2.5	Методы снижения размерности данных 1. Метод главных компонент 2. Факторный анализ	2			Метод главных компонент (4 ч)	Отчет по лабораторной работе
3	Методы классификации (4 ч)	4				
3.1	Деревья решений 1. Деревья решений 2. Анализ данных с использованием деревьев решений 3. Алгоритмы построения деревьев решений	2				
3.2	Методы классификации на основе байесовских сетей и расстояний 1. Байесовская классификация 2. Метод k-ближайших соседей 3. Метод наименьшего расстояния	2				Тест
4	Нейронные сети (10 ч)	2			8	
4.1	Нейронные сети 1. Нейронная сеть 2. Алгоритмы обучения нейронных сетей 3. Нейронные сети Кохонена				Нейронные сети. Слой Кохонена (4 ч)	Тест. Отчет по лабораторной работе
					Нейронные сети. Карты Кохонена (4 ч)	Отчет по лабораторной работе
5	Кластерный анализ (12 ч)	4			8	
5.1	Методы кластерного анализа 1. Кластерный анализ 2. Иерархические методы кластерного анализа	2			Иерархические методы кластерного анализа (4 ч)	Отчет по лабораторной работе
5.2	Неиерархические и графовые методы кластерного анализа 1. Неиерархические методы кластерного анализа 2. Графовые методы кластерного анализа	2			Неиерархические методы кластерного анализа (4 ч)	Тест. Отчет по лабораторной работе
6	Ассоциативные правила (2 ч)	2				
6.1	Методы поиска ассоциативных правил 1. Ассоциативные правила 2. Алгоритмы поиска ассоциативных правил	2				
7	Стохастические методы поиска (6 ч)	2			4	
7.1	Стохастические методы поиска 1. Методы случайного поиска 2. Эволюционные стратегии случайного поиска	2			Стохастические методы поиска (4ч)	Отчет по лабораторной работе

	3. Генетические алгоритмы					
8	Методы визуализации данных (2 ч)	2				
8.1	Методы визуализации данных 1. Визуальное представление данных 2. Методы визуализации многомерных данных 3. Принципы качественной визуализации	2				Тест
9	Процесс интеллектуального анализа данных (2 ч)	2				
9.1	Процесс интеллектуального анализа данных	2				
10	Стандарты и рынок инструментов интеллектуального анализа данных	4			6	
10.1	Стандарты и рынок инструментов интеллектуального анализа данных 1. Организационные и человеческие факторы 2. Стандарты ИАД 3. Рынок инструментов ИАД	2				
10.2	Интеллектуальный анализ данных с использованием пакета Matlab 1. Интерфейс Matlab 2. Основные элементы m-языка программирования 3. Анализ многомерных данных	2			Основы работы в Matlab (6 ч)	Тест. Отчет по лабораторной работе

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная и дополнительная литература

№№ п/п	Список литературы
	Основная
1.	Анализ данных и процессов: учеб. пособие / А. А. Барсегян [и др.]. – 3-е изд., перераб. и доп. – СПб., 2009. – 512 с.
2.	Афифи, А. Статистический анализ: Подход с использованием ЭВМ : пер. с англ. / А. Афифи, С.Эйзен. – М., 1982. – 488 с.
3.	Прикладная статистика: Классификация и снижение размерности : справ. изд. / С. А. Айвазян [и др.] ; под ред. С. А. Айвазяна. – М., 1989. – 607 с.
4.	Лагутин, М. Б. Наглядная математическая статистика : учеб. пособие / М. Б. Лагутин. – М., 2007. – 472 с.
5.	Яцков, Н. Н. Интеллектуальный анализ данных : пособие / Н. Н. Яцков. – Минск : БГУ, 2014. – 151 с.
	Дополнительная
1.	Айвазян, С. А. Прикладная статистика. Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин ; под ред. С. А. Айвазяна. – М., 1985. – 487 с.
2.	Дронов, С. В. Многомерный статистический анализ : учеб. пособие / С. В. Дронов. – Барнаул, 2003. – 213 с.
3.	Дюк, В. Data Mining: учебный курс / В. Дюк, А. Самойленко. –СПб., 2001. 368 с.
4.	Жиглявский, А. А. Методы поиска глобального экстремума / А. А. Жиглявский, А. Г. Жилинскас. – М., 1991. – 248 с.
5.	Ивченко, Г. И. Введение в математическую статистику : учебник / Г. И. Ивченко, Ю. И. Медведев. – М., 2010. – 600 с.
6.	Кобзарь, А. И. Прикладная математическая статистика : для инженеров и научных работников / А. И. Кобзарь. – М., 2006. – 816 с.
7.	Мандель, И.Д. Кластерный анализ / И.Д. Мандель –М., 1988. – 176 с.
8.	Медведев, В. С. Нейронные сети. MATLAB 6 / В. С. Медведев, В. Г. Потемкин ; под общ. ред. В. Г. Потемкина. – М., 2002. – 496 с.
9.	Наследов, А. Д. Математические методы психологического исследования. Анализ и интерпретация данных : учеб. пособие / А. Д. Наследов. – СПб., 2004. – 392 с.
10.	Сегаран, Т. Программируем коллективный разум : пер. с англ. / Т. Сегаран. – СПб., 2008. – 368с.

Перечень тем лабораторных работ:

1. Основы работы в Matlab (6 часов).
2. Нелинейный метод наименьших квадратов (4 часа).
3. Метод главных компонент (4 часа).
4. Иерархические методы кластерного анализа (4 часов).
5. Неиерархические методы кластерного анализа (4 часов).
6. Нейронные сети. Слой Кохонена (4 часа).
7. Нейронные сети. Карты Кохонена (4 часа).
8. Стохастические методы поиска (4 часа).

Методические рекомендации по организации самостоятельной работы студентов

Для организации самостоятельной работы студентов по учебной дисциплине рекомендуется использовать современные информационные технологии: разместить в сетевом доступе комплекс учебных и учебно-методических материалов (программа курса, методические указания к лабораторным занятиям, задания в тестовой форме, список рекомендуемой литературы и информационных ресурсов и др.).

Перечень используемых средств диагностики результатов учебной деятельности студентов

В качестве формы итогового контроля по дисциплине используется зачет. Оценка учебных достижений студента на зачете производится по десятибалльной шкале.

Для оценки уровня знаний студентов используется следующий диагностический инструментарий:

- компьютерные тесты по темам лекций;
- компьютерные тесты по темам лабораторных работ;
- проверка выполнения лабораторных работ и защита индивидуальных заданий при выполнении лабораторных работ;
- компьютерное итоговое тестирование.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УВО
на _____ / _____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 20__ г.)
(название кафедры)

Заведующий кафедрой

(степень, звание)

(подпись)

В.В. Скакун
(И.О.Фамилия)

УТВЕРЖДАЮ
Декан факультета

(степень, звание)

(подпись)

С.В. Малый
(И.О.Фамилия)