

Белорусский государственный университет

УТВЕРЖДАЮ
Проректор по учебной работе


А. Л. Голстик

2 мая 2017
Регистрационный № УД-3168/уч.

ВВЕДЕНИЕ В КОМПЬЮТЕРНЫЙ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Учебная программа учреждения высшего образования
по учебной дисциплине
для специальности:

1-31 03 04 Информатика

2017 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 03 04-2013, 1-98-01-01-1-2013, и учебных планов УВО № G31-169/уч. 2013 г. и УВО № G31-198/уч. 2013 г.

СОСТАВИТЕЛИ:

Казаченок Виктор Владимирович, профессор кафедры компьютерных технологий и систем Белорусского государственного университета, доктор педагогических наук, кандидат физико-математических наук, профессор

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой компьютерных технологий и систем Белорусского государственного университета
(протокол № 9 от 28 апреля 2017 г.)

Методической комиссией факультета прикладной математики и информатики Белорусского государственного университета
(протокол № 5 от 28 апреля 2017 г.)



Пояснительная записка

Дисциплина специализации "Введение в компьютерный и интеллектуальный анализ данных" входит в разряд дисциплин, читаемых студентам специальности "Информатика". Для успешного освоения дисциплины студентам понадобятся базовые навыки работы с приложениями Microsoft Office.

Дисциплина «Введение в компьютерный и интеллектуальный анализ данных» знакомит студентов с прикладными задачами и способами исследования зависимостей между различными данными и дальнейшем использовании знаний об этих зависимостях в задачах обработки и визуализации. Решение различных проблем зависит, прежде всего, от выявления взаимосвязи между различными явлениями, которые представляют собой результат одновременного воздействия большого числа причин. Поэтому при изучении этих явлений необходимо в первую очередь выявлять главные, основные причины. В то же время для получения достоверных выводов важно правильно оценить другие неучтенные и случайные причины. Решение перечисленных задач проводится, в первую очередь, методами прикладного статистического анализа, чем и обусловлено основное содержание дисциплины.

Цель дисциплины

Изучение теоретических основ предварительного (домодельного) статистического анализа данных (Exploratory Data Analysis); формирование навыков практического решения задач статистического анализа данных и представления получаемых результатов с использованием языка *R*.

Образовательная цель дисциплины

Формирование составной части банка знаний, получаемых будущими специалистами в процессе учебы и необходимых им в дальнейшем для успешной работы.

Развивающая цель дисциплины

Формирование у студентов основ математического мышления, изучение алгоритмов анализа данных при решении прикладных задач.

Основные задачи дисциплины

Основная задача - освоение студентами теории и практики анализа данных, анализ возможностей их визуализации и обработки.

Место в учебном процессе

Дисциплина "Введение в компьютерный и интеллектуальный анализ данных" изучается после получения базовых знаний по курсу «Программирование».

Требования к академическим компетенциям специалиста.

Специалист должен:

- АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач.
- АК-2. Владеть системным и сравнительным анализом.
- АК-3. Владеть исследовательскими навыками.
- АК-4. Уметь работать самостоятельно.
- АК-5. Быть способным порождать новые идеи (обладать креативностью).
- АК-6. Владеть междисциплинарным подходом при решении проблем.
- АК-7. Иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером.
- АК-9. Уметь учиться, повышать свою квалификацию в течение всей жизни.
- АК-10. Использовать основные законы естественно-научных дисциплин в профессиональной деятельности.
- АК-11. Владеть основными методами, способами и средствами получения, хранения, переработки информации с использованием компьютерной техники.
- СЛК-6. Уметь работать в команде.
- ПК-37. Пользоваться глобальными информационными ресурсами.

В результате изучения дисциплины студент должен:

знать:

- базовые математические модели анализа данных;
- основные математические методы анализа данных;
- различные математические алгоритмы анализа данных;
- основные математические методы визуализации;

уметь:

- выявлять взаимосвязи между различными явлениями;
- определять вид взаимосвязи между различными явлениями;
- проводить предварительную обработку информации;
- определять эффективные методы анализа данных;

владеть:

- основными методами анализа данных;
- основными методами визуализации данных.

Форма получения высшего образования: дневная (очная).

В соответствии с учебным планом по специальности 1-31 03 04 «Информатика», учебная программа предусматривает для изучения дисциплины 32 аудиторных часа, в том числе лекционных – 18 часов, лабораторных – 12 часов, УСР – 2 часа (2 курс, 3 семестр).

Форма текущей аттестации по учебной дисциплине – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

1. Первичная обработка статистических данных

Введение. Типы статистических данных и способы их первичной обработки. Качественные и количественные шкалы. Механический и случайный отбор при формировании выборки.

2. Числовые характеристики вариационных рядов

Методы предварительного анализа данных на основе дескриптивных статистик и графических представлений. Вариационные ряды, робастные показатели. Расчет средних значений, дисперсии, моды, медианы. Построение гистограмм, кумулянт частот, квартилей, определение квантилей.

3. Корреляционный анализ

Методы исследования парных статистических зависимостей на основе корреляционного анализа (для количественных, ранговых и номинальных шкал). Построение диаграммы рассеяния, корреляционной таблицы. Вычисление коэффициента корреляции, определение его значимости. Шкала Чеддока.

4. Регрессионный анализ

Методы исследования парных статистических зависимостей на основе регрессионного анализа. Метод наименьших квадратов. Вычисление неизвестных параметров линейных одномерных регрессионных моделей; графическое представление таких моделей. Матричное представление линейных регрессионных моделей.

5. Базовые конструкции языка *R*

Возможности и базовые конструкции языка *R*, предназначенные для реализации изучаемых разделов анализа данных (вычисление числовых характеристик вариационных рядов, корреляционный анализ, регрессионный анализ и др.). Форматы данных (векторы, матрицы, факторы, списки, пропущенные данные). Управляющие конструкции и функции. Статистические функции. Использование расширяющих библиотек и графических возможностей.

6. Ряды динамики

Методы фильтрации и прогнозирования рядов динамики. Составляющие временного ряда: тренд, циклические и случайные колебания. Индивидуальные и обобщающие характеристики. Методы анализа: а) средних, б) с использованием критерия Валлиса и Мура, в) с использованием критерия серий, г) скользящей средней, д) аналитического выравнивания.

7. Кластерный анализ

Алгоритмы кластерного анализа неоднородных многомерных данных. Стандартизация переменных. Характеристики кластера: центр, радиус, среднеквадратическое отклонение, размер. Агломеративные и дивизимные, иерархические и итеративные методы кластеризации. Алгоритмы кластеризации *K*-means и *G*-means.

8. Дискриминантный анализ

Методы и стадии интеллектуального анализа данных Data Mining (деревья решений, алгоритмы кластеризации, регрессионный анализ, нейронные сети, временные ряды и др.). Место интеллектуального анализа данных в процессе поддержки принятия решений. Общий вид искусственного нейрона, нейронная сеть Кохонена и ее геометрическая интерпретация.

9. Обучение нейронных сетей

Процесс Data Mining. Построение и использование модели. Прогнозирование на основании знания зависимостей, трендов и статистики. Место человека-аналитика в системах интеллектуального анализа данных. Обучение нейронных сетей для определения веса синаптических связей. Алгоритмы обучения с учителем и без учителя.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов				Количество часов УСР	Форма контроля знаний
		Лекции	Практические и семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8
1	Первичная обработка статистических данных	2		2			Лабораторная работа 1
2	Числовые характеристики вариационных рядов	2		2			
3	Корреляционный анализ	2		2			Лабораторная работа 2
4	Регрессионный анализ	2		2			Электронный тест, лабораторная работа 3
5	Базовые конструкции языка R	2					Лабораторная работа 4
6	Ряды динамики	2		2			
7	Кластерный анализ	2		2			Лабораторная работа 5
8	Дискриминантный анализ	2		1		1	Лабораторная работа 6
9	Обучение нейронных сетей	2		1		1	Электронный тест
И т о г о		18		14		2	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Рекомендуемая литература

№№ п/п	Список литературы	Год издания
	Основная	
1.	<i>Буяльская Ю.В.</i> Введение в компьютерный и интеллектуальный анализ данных/ Метод. указания / Ю.В. Буяльская, В.В. Казаченок – Минск: БГУ, 2016. – 46 с.	2016
2.	<i>Шипунов А.Б.</i> Наглядная статистика. Используем R. – М., 2012. – 116 с.	2012
3.	Шипунов А.Б., Балдин Е.М. Анализ данных с R. – М., 2013. – 224 с.	2013
4.	<i>Ефимова М.Р., Ганченко О.И., Петрова Е.В.</i> Практикум по общей теории статистики. М: Финансы и статистика, 2007. – 336 с.	2007
5.	<i>Таранчук В.Б.</i> Графический сервис вычислительного эксперимента : учеб.-метод. пособие. – Минск: БГУ, 2009. -124 с.	2009
6.	Теория статистики: Учебник/ Под ред. Р.А. Шмойловой. 4-е издание, доп. и перераб. – М.: Финансы и статистика, 2013. – 576 с.	2013
	Дополнительная	
1.	Прикладные системы обработки данных/ Метод. указания / И.С. Войтешенко, В.В. Казаченок, Т.В. Соболева, Т.В. Цеховая – Минск: БГУ, 2002. – 36 с.	2002
2.	Информатизация образования – 2012: материалы междунар. научн. конференции, Минск, 24–27 окт. 2012 г. / Белорусский гос. ун-т. – Минск: БГУ, 2012. – 368 с.	2012
3.	Информатизация образования – 2014: материалы междунар. научн. конференции, Минск, 22–25 окт. 2014 г. / Белорусский гос. ун-т. – Минск: БГУ, 2014. – 464 с.	2014

ПЕРЕЧЕНЬ ИСПОЛЬЗУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ РЕЗУЛЬТАТОВ УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ СТУДЕНТОВ

Для диагностики компетенций студентов используются следующие формы:

1. Устная форма.
2. Письменная форма.
3. Устно-письменная форма.
4. Техническая форма.

К устной форме диагностики компетенций относятся:

1. Собеседования.
2. Устные опросы.
3. Устный зачёт.

К письменной форме диагностики компетенций относятся:

1. Контрольные опросы.
2. Индивидуальные контрольные работы.
3. Письменные отчеты по лабораторным работам.
4. Оценивание на основе модульно-рейтинговой системы.

К устно-письменной форме диагностики компетенций относятся:

1. Отчеты по лабораторным работам с их устной защитой.
2. Оценивание на основе модульно-рейтинговой системы.

К технической форме диагностики компетенций относятся:

1. Электронные тесты.
2. Электронные отчёты по лабораторным работам.
3. Визуальные лабораторные работы.
4. Реализация алгоритма и результат выполнения программы.

ПЕРЕЧЕНЬ ЗАДАНИЙ УПРАВЛЯЕМОЙ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Тема 8. Дискриминантный анализ.

Задание 1. Составить программу, реализующую алгоритм кластеризации K-means.

Задание 2. Написать программу на языке R, реализующую алгоритм регрессионного анализа линейной зависимости.

Провести анализ качества построенной регрессионной функции.

Перечень используемых средств диагностики результатов управляемой самостоятельной работы студентов:

устный опрос, выполнение согласно заданиям индивидуальной работы, анализ индивидуальной работы.

Тема 9. Обучение нейронных сетей.

Задание 1. Реализовать простейший алгоритм обучения с учителем для определения весов синаптических связей.

Провести анализ качества обучения.

Перечень используемых средств диагностики результатов управляемой самостоятельной работы студентов:

подготовка электронной программы по индивидуальному заданию, индивидуальная защита учебного задания.

КОНТРОЛЬ КАЧЕСТВА УСВОЕНИЯ ЗНАНИЙ И ПРОВЕДЕНИЕ АТТЕСТАЦИИ

Контрольные мероприятия проводятся в соответствии с учебно-методической картой дисциплины. В случае неявки на контрольное мероприятие по уважительной причине студент вправе по согласованию с преподавателем выполнить его в дополнительное время. Для студентов, получивших неудовлетворительные оценки за контрольные мероприятия, либо не явившихся по неуважительной причине, по согласованию с преподавателем и с разрешения заведующего кафедрой мероприятие может быть проведено повторно.

Оценка текущей успеваемости рассчитывается на основе модульно-рейтинговой системы, основанной на Положении о рейтинговой системе БГУ от 18.08.2015 № 382-ОД.

Текущая аттестация предусматривает проведение зачёта и экзамена (в том числе с учётом результатов промежуточного и итогового тестирования) и проводится согласно правилам проведения аттестации студентов, курсантов, слушателей при освоении содержания образовательных программ высшего образования от 29 мая 2012 г. N 53.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола) ¹
Программирование нейронных сетей на языке Python	КТС		Протокол №9 от 28 апреля 2017г.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
 _____ (протокол № ____ от _____ 201_ г.)

 (название кафедры)

Заведующий кафедрой

 (ученая степень, ученое звание)
 (И.О.Фамилия)

 (подпись)

УТВЕРЖДАЮ

Декан факультета

 (ученая степень, ученое звание)
 (И.О.Фамилия)

 (подпись)