

Учреждение образования «Гомельский государственный технический университет имени П.О. Сухого»

УТВЕРЖДАЮ

Первый проректор ГГТУ им. П.О. Сухого


О.Д. Асенчик

28.06.2014

Регистрационный № УД 44-41 /уч.

ОСНОВЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности
1-40 05 01 «Информационные системы и технологии (по направлениям)»

Учебная программа составлена на основе образовательного стандарта высшего образования первой ступени ОСВО 1-40 05 01-2013; учебных планов учреждения образования «Гомельский государственный технический университет имени П.О. Сухого» специальности 1-40 05 01 «Информационные системы и технологии (по направлениям)», регистрационные №№ I 40-1-07/уч. 11.02.2016, I 40-1-13/уч. 23.05.2017, I 40-1-14/уч. 23.05.2017, I 40-1-30/уч. 17.02.2016, I 40-1-31/уч. 17.02.2016, I 40-1-15/уч. 23.05.2017.

СОСТАВИТЕЛЬ:

А.В. Цитринов, доцент кафедры «Информационные технологии» учреждения образования «Гомельский государственный технический университет имени П.О. Сухого», кандидат физико-математических наук, доцент.

РЕЦЕНЗЕНТ:

А.Н. Семенюта, заведующий кафедрой «Информационно-вычислительные системы» учреждения образования «Белорусский торгово-экономический университет потребительской кооперации», доктор технических наук, профессор

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой «Информационные технологии» учреждения образования «Гомельский государственный технический университет имени П.О. Сухого» (протокол № 17 от 08.05.2017 г.);

Научно-методическим советом факультета автоматизированных и информационных систем учреждения образования «Гомельский государственный технический университет имени П.О. Сухого» (протокол № 10 от 24.05.2017 г.); *УДП-04-33/ч 2.*

Научно-методическим советом заочного факультета учреждения образования «Гомельский государственный технический университет имени П.О. Сухого» (протокол № 5 от 1.06. 2017 г.); *УДП-123-16ч*

Научно-методическим Советом учреждения образования «Гомельский государственный технический университет имени П.О. Сухого» (протокол № 6 от 27.06. 2017 г.).

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная программа дисциплины «Основы интеллектуального анализа данных» разработана для студентов специальности 1-40 05 01 «Информационные системы и технологии (по направлениям)» направления 1-40 05 01-01 «Информационные системы и технологии (в проектировании и производстве)» для дневной, заочной и заочной сокращенной форм обучения и направления 1-40 05 01-012 «Информационные системы и технологии (в игровой индустрии)» для дневной формы получения образования.

Целью учебной дисциплины является изучение теоретических основ технологии интеллектуального анализа данных (ИАД), включая методы статистического анализа, классификации, нейронных сетей, 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 «Математика», «Теория вероятностей и математическая статистика», «Основы алгоритмизации и программирования», «Базы данных».

В свою очередь, учебная дисциплина «Основы интеллектуального анализа данных» является базовой для изучения учебной дисциплины «Оптимизация проектных решений».

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 для решения практических задач управления и обработки информации,
- творчески и эффективно использовать полученные знания в профессиональной деятельности.

Изучение учебной дисциплины «Основы интеллектуального анализа данных» должно обеспечить формирование у студента следующих компетенций:

АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач.

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-10. Использовать основные законы естественнонаучных дисциплин в профессиональной деятельности.

АК-11. Владеть основными методами, способами и средствами получения, хранения, переработки информации с использованием компьютерной техники.

АК-14. На научной основе организовывать свой труд, самостоятельно оценивать результаты своей деятельности.

ПК-4. Разрабатывать программные средства и системы обеспечения автоматизированной поддержки решений задач профессиональной деятельности.

ПК-5. Осуществлять контроль эффективности использования вычислительных средств и информационных систем в профессиональной деятельности.

ПК-7. Использовать методы анализа и мониторинга для приведения процессов профессиональной деятельности в соответствие действующим стандартам, инструкциям, правилам и нормам.

ПК-9. Выполнять моделирование и проектирование программных средств, разрабатываемых для обеспечения профессиональной деятельности.

ПК-10. Разрабатывать техническую и проектную документацию на создаваемые программные средства решений профессиональных задач.

ПК-11. Разрабатывать функциональные, информационные и другие модели формализованного представления процессов профессиональной деятельности.

ПК-21. Анализировать и оценивать собранные данные.

ПК-24. Пользоваться глобальными информационными ресурсами.

Методика преподавания дисциплины строится на сочетании лекций, лабораторных занятий и самостоятельной работы. Изучение каждой темы, помимо приведенных в учебной программе литературных источников, предполагает использование информационных ресурсов Internet.

Для оценки достижений студентов рекомендуется использовать следующий диагностический инструментарий:

- проведение текущих контрольных опросов по отдельным темам курса;
- письменные отчеты по лабораторным работам с их устной защитой;
- выступление студентов с докладами на студенческих научно-практических конференциях;
- сдача экзамена по дисциплине.

Форма получения высшего образования:

- по направлению 1-40 05 01-01 «Информационные системы и технологии (в проектировании и производстве)» - дневная, заочная, заочная на базе среднего специального образования;
- по направлению 1-40 05 01-012 «Информационные системы и технологии (в игровой индустрии)» - дневная.

Общее количество часов, количество аудиторных часов, трудоемкость учебной дисциплины.

Учебная программа дисциплины «Основы интеллектуального анализа данных» рассчитана на 90/114 часов всего. Аудиторных часов по дневной форме получения образования – 48, по заочной – 10, по заочной на базе среднего специального образования – 10. Трудоемкость учебной дисциплины составляет 2,5/3 зачетные единицы.

Распределение аудиторного времени по видам занятий, курсам и семестрам:

	ДО	ЗО	ЗОс
Курс	3	1,2/3	2
Семестр	5	2,3/5,6	3,4
Лекции (часов)	24	6	6
Лабораторные занятия (часов)	24	4	4
Всего аудиторных (часов)	48	10	10

Формы текущей аттестации по учебной дисциплине

Экзамен	/5	/6	4
Зачет	5/	3/	–

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.

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 – перцептрон. 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 методы 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 FPG.

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. 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. Визуальное представление данных. Анализ данных. Интерпретация и прогнозирование.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
(Дневная форма получения образования)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов СР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
Раздел 1. Введение в интеллектуальный анализ данных								
1.1	Основные понятия дисциплины «Основы интеллектуального анализа данных»	2						Э
Раздел 2. Статистические методы								
2.1	Методы предварительного анализа данных. Корреляционный анализ	2						Э
2.2	Регрессионный анализ. Дисперсионный анализ	2			4			О, ЗЛР, Э
2.3	Методы снижения размерности данных	2			4			О, ЗЛР, Э
Раздел 3. Методы классификации								
3.1	Деревья решений	2						О, ЗЛР, Э
3.2	Методы классификации на основе байесовских сетей и расстояний	2						О, ЗЛР, Э
Раздел 4. Нейронные сети								
4.1	Сети классификации и кластеризации данных. Архитектура нейронных сетей.	2						Э
4.2	Алгоритмы обучения нейронных сетей	2			4			О, ЗЛР, Э
Раздел 5. Кластерный анализ								
5.1	Методы кластерного анализа. Неиерархические и графовые методы кластерного анализа	2			4			О, ЗЛР, Э
Раздел 6. Ассоциативные правила								
6.1	Методы поиска ассоциативных правил	2						Э
Раздел 7. Стохастические методы поиска								
	Стохастические методы поиска	2			4			О, ЗЛР, Э
Раздел 8. Процесс интеллектуального анализа данных								
	Процесс интеллектуального анализа данных	2			4			О, ЗЛР, Э
	Итого	24			24			

Принятые обозначения: О – отчет по лабор. работе, ЗЛР – защита лабор. работы, Э – экзамен

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
(Заочная/заочная сокращенная форма получения образования)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов СР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
Раздел 1. Введение в интеллектуальный анализ данных								
1.1	Основные понятия дисциплины «Основы интеллектуального анализа данных»	0,5						Э
Раздел 2. Статистические методы								
2.1	Методы предварительного анализа данных. Корреляционный анализ	0,5						Э
2.2	Регрессионный анализ. Дисперсионный анализ	0,5						Э
2.3	Методы снижения размерности данных	0,5						Э
Раздел 3. Методы классификации								
3.1	Деревья решений	0,5						Э
3.2	Методы классификации на основе байесовских сетей и расстояний	0,5						Э
Раздел 4. Нейронные сети								
4.1	Сети классификации и кластеризации данных. Архитектура нейронных сетей.	0,5						Э
4.2	Алгоритмы обучения нейронных сетей	0,5			1			О, ЗЛР, Э
Раздел 5. Кластерный анализ								
5.1	Методы кластерного анализа. Неиерархические и графовые методы кластерного анализа	0,5			1			О, ЗЛР, Э
Раздел 6. Ассоциативные правила								
6.1	Методы поиска ассоциативных правил	0,5						Э
Раздел 7. Стохастические методы поиска								
	Стохастические методы поиска	0,5						Э
Раздел 8. Процесс интеллектуального анализа данных								
	Процесс интеллектуального анализа данных	0,5			2			О, ЗЛР, Э
	Итого	6 ✓			4 ✓			

Принятые обозначения: О – отчет по лабор. работе, ЗЛР – защита лабор. работы, Э – экзамен.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

1. Макленнен, Дж. Microsoft SQL Server 2008. Data Mining – интеллектуальный анализ данных : перевод с английского / Д. Макленнен, Ч. Танг, Б. Криват. - Санкт-Петербург : БХВ-Петербург, 2009. - 700 с.
2. Кулаичев А. П. Методы и средства комплексного анализа данных : учеб. пособие для вузов. - 4-е изд., перераб. и доп.. - Москва : ФОРУМ : ИНФРА-М, 2006. - 511 с.
3. Прикладная статистика: Классификация и снижение размерности: Справочное изд. / Под ред. С.А. Айвазяна. - М. : Финансы и статистика, 1989. - 607с.
4. Литтл Р. Дж. Статистический анализ данных с пропусками. - Москва : Финансы и статистика, 1991. - 333 с.
5. Петрунин, Ю. Ю. Информационные технологии анализа данных. Data Analysis : учебное пособие / Ю. Ю. Петрунин. - Москва : Книжный дом-университет, 2008. - 291 с.
6. Чашкин, Ю. Р. Математическая статистика. Анализ и обработка данных : учебное пособие для вузов / Ю. Р. Чашкин. - Изд. 2-е, перераб. и доп. - Ростов-на-Дону : Феникс, 2010. - 236, [1] с.

Дополнительная литература

7. Анализ данных и процессов: учеб. пособие / А. А. Барсегян [и др.]. – 3-е изд., перераб. и доп. – СПб., 2009. – 512 с.
8. Яцков, Н. Н. Интеллектуальный анализ данных : пособие / Н. Н. Яцков. – Минск : БГУ, 2014. – 151 с.
9. Ивченко, Г. И. Введение в математическую статистику : учебник / Г. И. Ивченко, Ю. И. Медведев. – М., 2010. – 600 с.
10. Каллан, Р. Основные концепции нейронных сетей / Р. Каллан. - Москва : Вильямс, 2003. - 287 с.
11. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / пер. с польск. И. Д. Рудинского. - Москва : Горячая линия-Телеком, 2008. - 383 с.
12. O'Neil, C. Doing Data Science Straight Talk from the Frontline / C. O'Neil, R. Schutt – Sebastopol : O'Reilly Media, 2013. – 408 p.
13. McKinney, W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython / W. McKinney – Sebastopol : O'Reilly Media, 2014. – 452 p.

Список литературы взят с сайта АИИ (Писцова И.В.)

Примерный перечень тем лабораторных работ

1. Нелинейный метод наименьших квадратов.
2. Метод главных компонент.
3. Иерархические методы кластерного анализа.
4. Неиерархические методы кластерного анализа.
5. Нейронные сети. Слой и карты Кохонена.
6. Стохастические методы поиска.

Методы и технологии обучения

Основные рекомендуемые методы и технологии обучения, отвечающие цели и задачам дисциплины:

- проблемное обучение (вариативное изложение, проблемное изложение, частично-поисковый метод), реализуемое на лекционных занятиях;
- учебно-исследовательская деятельность, творческий подход, реализуемые на лабораторных занятиях.

Организация самостоятельной работы студентов

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы: подготовка к выполнению и защите лабораторных работ; подготовка к экзамену/зачету.

Диагностика компетенций студента

Учебной программой по специальности 1-40 05 01 «Информационные системы и технологии (по направлениям)», направление специальности 1-40 05 01-01 «Информационные системы и технологии (в проектировании и производстве)» в качестве формы текущей аттестации по учебной дисциплине «Основы интеллектуального анализа данных» предусмотрен экзамен.

Для промежуточного контроля по учебной дисциплине и диагностики компетенций студентов могут использоваться следующие формы:

- собеседование;
- защита лабораторных работ;
- устный опрос;
- проведение текущих опросов по отдельным темам дисциплины;
- выступление студента по разработанной им теме;
- решение проблемных задач.

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)
Оптимизация проектных решений	ИТ	Нет	Протокол № 17 от 08.05.2017 г.

Заведующий кафедрой
«Информационные технологии»,
к.т.н., доцент



К.С. Курочка

Библиотека ГГТУ ИМ Л.Ф.Щеголева