

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Ректор Белорусского
государственного университета

А.Д.Король

10 июня 2024 г.

Регистрационный № 1849/м.



ПРОГРАММНЫЕ СРЕДСТВА АНАЛИЗА ДАННЫХ

Учебная программа учреждения образования
по учебной дисциплине для специальности:

7-06-0533-05 Прикладная математика и информатика

Профилизации:

Компьютерный анализ данных

Интеллектуальные системы

2024 г.

Учебная программа составлена на основе ОСВО 7-06-0533-05-2023 и учебных планов № М53-5.3-43/уч. от 15.02.2023, № М53-5.3-79/уч. от 11.04.2023.

СОСТАВИТЕЛЬ:

П.А. Пашук, старший преподаватель кафедры теории вероятностей и математической статистики факультета прикладной математики и информатики Белорусского государственного университета, магистр прикладной математики и информационных технологий.

РЕЦЕНЗЕНТЫ:

М.С. Абрамович, заведующий НИЛ статистического анализа и моделирования учреждения Белорусского государственного университета «НИИ прикладных проблем математики и информатики», кандидат физико-математических наук, доцент;

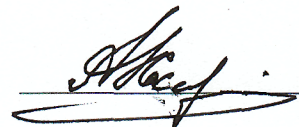
А.И. Кишкар, ведущий инженер-программист отдела информационных систем управления бизнес-приложений департамента производства ЗАО «Международный деловой альянс», магистр прикладной математики и информационных технологий.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики БГУ
(протокол № 12 от 21.05.2024);

Научно-методическим советом БГУ
(протокол № 8 от 31.05.2024)

Заведующий кафедрой теории вероятностей
и математической статистики



А.Ю.Харин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины «Программные средства анализа данных» – формирование необходимых навыков программной реализации методов и алгоритмов, а также их применения при анализе данных.

Задачи учебной дисциплины:

- 1) изучение базового синтаксиса и особенностей программной среды при анализе данных;
- 2) знакомство студентов с применением методов программной среды, а также их преимуществ и недостатками;
- 3) формирование практических навыков решения прикладных задач с использованием современного программного обеспечения.

Место учебной дисциплины в системе подготовки специалиста с углубленным высшим образованием.

Учебная дисциплина относится к модулю «Специальные программные средства» государственного компонента.

Учебная программа составлена с учетом межпредметных **связей** и программ по дисциплинам: «Многомерный статистический анализ», «Методы визуализации в анализе данных средствами языка R», «Методы статистического анализа сложных данных», «Математические модели и методы компьютерного зрения».

Требования к компетенциям

Освоение учебной дисциплины «Программные средства анализа данных» должно обеспечить формирование следующих компетенций:

универсальные компетенции:

УК. Осуществлять коммуникации на иностранном языке в академической, научной и профессиональной среде для реализации научно-исследовательской и инновационной деятельности;

углубленные профессиональные компетенции:

УПК. Применять знания по современным вероятностным моделям, используемым для анализа сложных данных, применять специальные современные методы анализа сложных данных.

В результате освоения учебной дисциплины студент должен:

знать:

- базовый синтаксис и основные стандартные модуля изучаемой программной среды для анализа данных;
- особенности программной среды при реализации методов и алгоритмов анализа данных;

уметь:

- реализовывать методы и алгоритмы анализа данных в программной среде;
- выбирать оптимальный способ реализации методов преобразования и анализа данных;

иметь навык:

- работы с основными методами анализа данных в программной среде;
- компьютерной реализации основных методов;
- решения прикладных задач с использованием современного программного обеспечения;

Структура учебной дисциплины

Дисциплина изучается в 1 семестре. Всего на изучение учебной дисциплины «Программные средства анализа данных» отведено:

- для очной формы получения углубленного высшего образования: 90 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Введение в язык программирования.

Основные сведения о языке. Версии языка. Основные используемые среды разработки. Общее описание процесса разработки. Настройка рабочей среды.

Тема 2. Стандартные типы данных.

Основные типы данных. Векторные и матричные структуры для хранения данных. Списки и таблицы данных. Доступ к элементам списка или таблицы

Тема 3. Обработка данных.

Векторные вычисления. Принцип циклического повторения. Основные операции. Операции над матрицами. Сортировка. Работа с пропущенными значениями.

Тема 4. Условные операторы выбора и операторы циклов.

Условный оператор выбора. Условный оператор множественного выбора. Операторы циклов `for` и `while`.

Тема 5. Функции.

Создание новых функций, определенных пользователем. Область видимости переменных. Именованные аргументы, значения по умолчанию, произвольное количество аргументов. Функция в качестве аргумента функции. Функция, возвращающая функцию. Рекурсия.

Тема 6. Организация ввода-вывода данных.

Редактор данных. Доступ к внешним данным. Чтение данных с клавиатуры. Чтение данных из внешнего файла, страницы в сети Интернет. Сохранение данных в файл.

Тема 7. Основные библиотеки анализа данных.

Вычисление базовых описательных статистик. Проверка гипотез о принадлежности выборки заданному распределению. Построение простейших графиков. Построение классические графиков (гистограмм, ящиков с усами, графиков рассеяния и пр.). Точная настройка атрибутов и свойств графиков. Генерация дискретных и непрерывных случайных величин. Метод Монте-Карло. Кластеризация и дискриминантный анализ. Применение алгоритмов машинного обучения.

Тема 8. Подготовка отчетов.

Особенности подготовки отчета в среде LaTeX, содержащего скрипты и вставки языка. Работа с документами в разных кодировках.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения высшего образования с применением дистанционных образовательных технологий (ДОТ)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Формы контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	Введение в язык программирования	2			2			Устный опрос
2	Стандартные типы данных	2			2			Индивидуальные задания
3	Обработка данных	2			2			Расчетно-графическая работа, отчет по лабораторной работе
4	Условные операторы выбора и операторы циклов	2			2			Контрольная работа
5	Функции	2			2			Индивидуальные задания, отчет по лабораторной работе
6	Организация ввода-вывода данных	2			2			Контрольная работа
7	Основные библиотеки анализа данных	6			6			Индивидуальные задания, контрольная работа
8	Подготовка отчетов	2			2			Отчет по лабораторной работе
ИТОГО		20			20			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

1. Васильев, Ю. Python для Data Science / Ю. Васильев; [пер. с англ. А. Алимова]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 270 с. – URL: <https://ibooks.ru/bookshelf/390133/reading>.
2. Брюс, П. Практическая статистика для специалистов Data Science: 50+ важнейших понятий с использованием R и Python / П. Брюс, Э. Брюс, П. Гедек; [пер. с англ. А. Логунова]. – 2-е изд., перераб. и доп. – Санкт-Петербург: БХВ-Петербург, 2021. – 346 с. – URL: <https://ibooks.ru/reading.php?short=1&productid=380029>.
3. Ланц, Б. Машинное обучение на R: экспертные техники для прогностического анализа / Б. Ланц; [пер. с англ. Е. Сандицкой]. – Санкт-Петербург [и др.]: Питер, 2020. – 462 с. – URL: <https://ibooks.ru/bookshelf/367984>.
4. Яворски, М. Python. Лучшие практики и инструменты / М. Яворски, Т. Зиаде; [пер. с англ. Е. Матвеев]. – 3-е изд. – Санкт-Петербург; Москва; Минск: Питер, 2021. – 558 с. – URL: <https://ibooks.ru/bookshelf/376831>.
5. Доусон, М. Програмируем на Python / М. Доусон; [пер. с англ. В. Порицкий]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 414 с.

Дополнительная литература

1. Мэтлофф, Н. Искусство программирования на R. Погружение в большие данные / Н. Мэтлофф. – Санкт-Петербург [и др.]: Питер, 2019. – 413 с.: ил.
2. Мастицкий, С.Э. Статистический анализ и визуализация данных с помощью R / Мастицкий С. Э., Шитиков В. К. – Москва: ДМК Пресс, 2015. – 495 с.: ил.
3. Хобсон, Л. Обработка естественного языка в действии / Л. Хобсон, Х. Ханнес, Х. Коул; [пер. с англ. И. Пальти, С. Черников]. – Санкт-Петербург [и др.]: Питер, 2020. – 575 с.: ил.
4. Кабаков, Р.И. R в действии / Р.И. Кабаков. – 3-е изд. – Москва: ДМК Пресс, 2023. – 768 с.
5. Бейдер, Д. Знакомство с Python / Д. Бейдер и др.; [пер. с англ. Е. Матвеев]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 507 с.: ил.
6. Вандер Плас, Дж. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас; [пер. с англ. И. Пальти]. – Санкт-Петербург; Москва; Минск: Питер, 2023. – 573 с.: ил.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Объектом диагностики компетенций студентов являются знания, умения, полученные ими в результате изучения учебной дисциплины. Выявление учебных достижений студентов осуществляется с помощью мероприятий текущего контроля и промежуточной аттестации.

Для диагностики компетенций могут использоваться следующие средства текущего контроля: устный опрос, контрольные работы, отчеты по лабораторным работам, индивидуальные задания, расчётно-графические работы.

Формой промежуточной аттестации по дисциплине «Программные средства анализа данных» учебным планом предусмотрен **экзамен**.

Для формирования итоговой отметки по учебной дисциплине используется модульно-рейтинговая система оценки знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая система предусматривает использование весовых коэффициентов для текущей и промежуточной аттестации студентов по учебной дисциплине.

Формирование итоговой отметки в ходе проведения контрольных мероприятий текущей аттестации (примерные весовые коэффициенты, определяющие вклад текущей аттестации в отметку при прохождении промежуточной аттестации):

- контрольные работы – 40%;
- защита отчетов по лабораторным работам – 60%.

Экзамен по дисциплине проходит в устной и(или) письменной форме. В случае успешного прохождения всех форм текущего контроля и получения положительной (4 и выше) отметки текущей аттестации, допускается получение экзамена без проведения дополнительного опроса. При этом явка обучающегося на экзамен является обязательной.

Итоговая отметка по дисциплине рассчитывается на основе итоговой отметки текущей аттестации (рейтинговой системы оценки знаний) 60% и отметки на экзамене 40%.

Примерная тематика лабораторных занятий

Занятие № 1. Введение в язык программирования.

Занятие № 2. Стандартные типы данных.

Занятие № 3. Обработка данных.

Занятие № 4. Условные операторы выбора и операторы циклов.

Занятие № 5. Функции.

Занятие № 6. Организация ввода-вывода данных.

Занятие № 7. Вычисление описательных статистик. Проверка простых гипотез.

Занятие № 8. Моделирование случайных величин. Метод Монте-Карло.

Занятие № 9. Кластеризация и дискриминантный анализ.

Занятие № 10. Подготовка отчетов.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется *практико-ориентированный подход*, который предполагает:

- освоение содержания образования через решения практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
- использованию процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

Методические рекомендации по организации самостоятельной работы обучающихся

Самостоятельная работа с целью изучения материала учебной дисциплины предполагает работу с рекомендованной учебной литературой и Интернет-ресурсами. Теоретические сведения закрепляются выполнением лабораторных заданий, при выполнении которых следует руководствоваться методическими разработками, размещенными в электронной библиотеке университета и на образовательном портале. Также могут быть предложены дополнительные задания (тесты, задания для самостоятельного выполнения) для самооценивания и более глубокого усвоения полученного материала.

Примерный перечень вопросов к экзамену

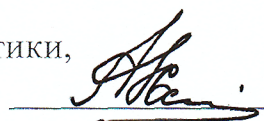
1. Базовые типы данных и операции с ними.
2. Структуры данных и их особенности.
3. Логические выражения.
4. Индексация и обращение к элементам массивов.
5. Условные конструкции.
6. Циклы.
7. Базовый синтаксис создания функций.
8. Создание функций с бесконечным числом аргументов.
9. Создание функций, с заданными по умолчанию аргументами.
10. Примеры создания функций, возвращающих функцию.
11. Примеры создания функции, использующих функции в качестве аргумента.
12. Глобальное и локальное окружения. Суперприсваивание.
13. Подключение и использование дополнительных пакетов для анализа данных.
14. Импорт и экспорт данных.
15. Графическое представление данных.

16. Моделирование случайных величин.
17. Проверка статистических гипотез.
18. Тесты согласия.
19. Корреляционный анализ.
20. Построение моделей простой линейной регрессии.
21. Модели нелинейной регрессии.
22. Анализ остатков в моделях регрессии.
23. Создание отчетов.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

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)
Многомерный статистический анализ	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Методы визуализации в анализе данных средствами языка R	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Методы статистического анализа сложных данных	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024
Математические модели и методы компьютерного зрения	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 12 от 21.05.2024

Заведующий кафедрой теории вероятностей и математической статистики,
доктор физ.-мат. наук, профессор



А.Ю. Харин

21 мая 2024 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ

Декан факультета
