

Государственное образовательное учреждение
«Приднестровский государственный университет им. Т.Г. Шевченко»

Физико-технический институт
Физико-математический факультет
Кафедра высшей и прикладной математики и информатики

УТВЕРЖДАЮ

Директор физико-технического института

Д.Н. Калошин

« » 2024 г.



РАБОЧАЯ ПРОГРАММА

по дисциплине

Б1.В.07 «Введение в компьютерный анализ данных»

на 2024/2025 учебный год

Направление

01.03.02 «Прикладная математика и информатика»

Профиль

Системное программирование и компьютерные технологии

Квалификация

Бакалавр

Форма обучения

Очная

ГОД НАБОРА 2022

Тирасполь 2024 г.

Рабочая программа дисциплины «Введение в компьютерный анализ данных» разработана в соответствии с требованиями Государственного образовательного стандарта ВО по направлению подготовки 01.03.02 «Прикладная математика и информатика» и основной профессиональной образовательной программы (учебного плана) по профилю подготовки «Системное программирование и компьютерные технологии».

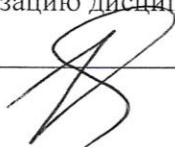
Составитель рабочей программы:

ст. преподаватель  Калинкова Е.В.

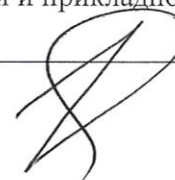
Рабочая программа утверждена на заседании кафедры высшей и прикладной математики и информатики

« 14 » 11 2024 г. протокол № 4

Зав. кафедрой, отвечающей за реализацию дисциплины

« 14 » 11 2024 г.  Коровай А.В.

Зав. выпускающей кафедрой высшей и прикладной математики и информатики

« 14 » 11 2024 г.  Коровай А.В.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Введение в компьютерный анализ данных» является формирование у студентов теоретических знаний, практических умений и навыков по применению методов анализа данных с использованием современных программных средств.

Задачами освоения дисциплины «Введение в компьютерный анализ данных» являются:

- овладение теоретическими основами наиболее распространенных методов анализа данных и условиями их применения;
- изучение основных методов поиска в данных внутренних закономерностей, взаимосвязей, тенденций;
- изучение основ количественных методов оценки адекватности и точности построенных моделей;
- изучение принципов работы программных средств, предназначенных для решения задач анализа данных;
- формирование практических навыков применения стандартных и универсальных программных средств, применяемых для анализа данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина Б1.В.07 «Введение в компьютерный анализ данных» относится к дисциплинам части, формируемой участниками образовательных отношений, Блока 1.

Для освоения дисциплины «Введение в компьютерный анализ данных» обучающиеся используют знания, умения и навыки, сформированные в ходе изучения дисциплин «Теория вероятностей и математическая статистика», «Языки и методы программирования (Практикум на ЭВМ)».

Дисциплина «Введение в компьютерный анализ данных» является предшествующей для изучения дисциплины «Интеллектуальный анализ данных».

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Изучение дисциплины направлено на формирование компетенций, приведенных в таблице ниже

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения компетенции
Универсальные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		
Общепрофессиональные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		
Обязательные профессиональные компетенции и индикаторы их достижения		
	ПК-2 Способен понимать и применять в исследовательской и прикладной деятельности современный математический аппарат.	ИД-1ПК-2 Знает современный математический аппарат.
		ИД-2ПК-2 Умеет применять методы, алгоритмы и приёмы современного математического аппарата.
		ИД-3ПК-2 Владеет практическими навыками применения современного математического аппарата в исследовательской и прикладной деятельности.
	ПК-4 Способен демонстрировать знания современных языков	ИД-1ПК-4 Знает основные языки программирования и основы работы с базами данных, операционные системы и оболочки,

	программирования, операционных систем, офисных приложений, информационно-телекоммуникационной сети "Интернет", способов и механизмов управления данными, принципов организации, состава и схемы работы операционных систем.	современные программные среды разработки информационных систем и технологий.
		ИД-2ПК-4 Умеет применять языки программирования, современные программные среды разработки информационных систем и технологий для автоматизации бизнес-процессов, решения прикладных задач различных классов, ведения баз данных и информационных хранилищ.
		ИД-3ПК-4 Владеет навыками программирования, отладки и тестирования прототипов программно-технических комплексов задач.
Рекомендуемые профессиональные компетенции и индикаторы их достижения		
Учебным планом не предусмотрены		

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Распределение трудоемкости в з.е./часах по видам аудиторной и самостоятельной работы студентов по семестрам

Семестр	Трудоемкость, з.е./часы	Количество часов					Форма контроля
		В том числе					
		Аудиторных				Самостоятельная работа (СР)	
		Всего	Лекций (Л)	Практических занятий (ПЗ)	Лабораторных занятий (ЛЗ)		
6	3/108	54	18		36	54	Зачет с оценкой
Итого:	3/108	54	18		36	54	

4.2. Распределение видов учебной работы и их трудоемкости по разделам дисциплины

№ раздела	Наименование разделов	Всего	Аудиторная работа			СР
			Л	ПЗ	ЛЗ	
1	Введение в анализ данных.	39	6		14	19
2	Корреляционный анализ. Регрессионный анализ.	24	4		8	12
3	Классификация и кластеризация.	20	4		6	10
4	Анализ и прогнозирование временных рядов.	13	2		4	7
5	Другие методы анализа.	12	2		4	6
Итого:		108	18		36	54

4.3. Тематический план по видам учебной деятельности

Лекции

№ п/п	Номер раздела дисциплины	Объем часов	Тема лекции	Учебно-наглядные пособия
Введение в анализ данных				
1	1	2	Понятие анализа данных. Основные задачи анализа данных. Критерии классификации статистических данных. Шкалы измерений.	презентация
2		2	Генеральная совокупность и выборка из нее. Вариационные ряды. Описательная статистика.	презентация
3		2	Точечные и интервальные оценки параметров распределения. Проверка статистических гипотез. Анализ двух выборок.	презентация
Итого по разделу часов:		6		
Корреляционный анализ. Регрессионный анализ.				
4	2	2	Понятие и свойства корреляционной связи. Коэффициент линейной корреляции Пирсона. Проверка значимости линейного коэффициента корреляции.	презентация
5		2	Регрессионный анализ и его задачи. Регрессионные модели. Линейная регрессия. Множественная регрессия. Проверка значимости уравнения и коэффициентов регрессии.	презентация
Итого по разделу часов:		4		
Классификация и кластеризация				
6	3	2	Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Наивный байесовский классификатор.	презентация
7		2	Постановка задачи кластерного анализа. Иерархические и неиерархические методы кластеризации. Метод k-средних.	презентация
Итого по разделу часов:		4		
Анализ и прогнозирование временных рядов				
8	4	2	Временные ряды и их характеристики. Аддитивная и мультипликативная модель. Метод скользящего среднего. Метод экспоненциального сглаживания.	презентация
Итого по разделу часов:		2		
Другие методы анализа				
9	5	2	Понятие, назначение и виды дисперсионного анализа. Однофакторный дисперсионный анализ. Факторный анализ.	презентация
Итого по разделу часов:		2		
Итого:		18		

Лабораторные занятия

№ п/п	Номер раздела дисциплины	Объем часов	Тема лабораторных занятий	Учебно-наглядные пособия
Введение в анализ данных				
1	1	2	Построение распределений случайных величин.	лаб. практикум
2		4	Построение вариационных рядов.	
3		2	Вычисление выборочных характеристик данных.	лаб. практикум
4		2	Анализ одной выборки.	лаб. практикум
5		2	Проверка гипотезы о распределении случайной величины.	
6		2	Параметрические методы анализа двух выборок.	лаб. практикум
Итого по разделу часов:		14		
Корреляционный анализ. Регрессионный анализ.				
7	2	4	Корреляционный анализ.	лаб. практикум
8		4	Регрессионный анализ.	лаб. практикум
Итого по разделу часов:		8		
Классификация и кластеризация				
9	3	2	Наивный байесовский классификатор.	лаб. практикум
10		4	Кластерный анализ.	лаб. практикум
Итого по разделу часов:		6		
Анализ и прогнозирование временных рядов				
11	4	4	Анализ и прогнозирование временных рядов.	лаб. практикум
Итого по разделу часов:		4		
Другие методы анализа				
12	5	2	Дисперсионный анализ.	лаб. практикум
13		2	Факторный анализ.	лаб. практикум
Итого по разделу часов:		4		
Итого:		36		

Самостоятельная работа обучающегося

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
Раздел 1	1	Понятие анализа данных. Основные задачи анализа данных. Программное обеспечение анализа данных. (1,2)	2
	2	Критерии классификации статистических данных. Шкалы измерений. (1,2)	2
	3	Генеральная совокупность и выборка из нее. Вариационные ряды. Графическое представление вариационных рядов. (1,2,3)	3
	4	Описательная статистика. (1,2,3)	2
	5	Точечные и интервальные оценки параметров распределения. (1,2,3)	2
	6	Общая методика проверки гипотез. Проверка гипотез о законе распределения. (1,2,3)	3
	7	Параметрические критерии проверки гипотез. Проверка	3

		гипотезы о равенстве двух выборочных средних. Проверка гипотезы о равенстве двух выборочных дисперсий. (1,2,3)	
	8	Непараметрические критерии проверки гипотез. (1,2)	2
Итого по разделу часов:			19
Раздел 2	1	Понятие корреляционной связи. Свойства корреляционной связи. (1,2,3)	2
	2	Коэффициент линейной корреляции Пирсона. Проверка значимости линейного коэффициента корреляции. (1,2,3)	2
	3	Регрессионный анализ и его задачи. Регрессионные модели. Метод наименьших квадратов. (1,2,3)	2
	4	Двумерная линейная модель регрессии. Проверка значимости уравнения и коэффициентов регрессии. (1,2,3)	2
	5	Множественная линейная модель регрессии. (1,2,3)	2
	6	Нелинейные модели регрессии. (1,2,3)	2
Итого по разделу часов:			12
Раздел 3	1	Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. (1,2)	2
	2	Формула Байеса. Наивный байесовский классификатор. (1,2,3)	2
	3	Основные понятия и определения кластерного анализа. Классификация методов кластеризации. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. (1,2)	2
	4	Иерархические методы кластеризации. Агломеративная кластеризация. (1,2,3)	2
	5	Неиерархические методы кластеризации. Метод k -средних. (1,2,3)	2
Итого по разделу часов:			10
Раздел 4	1	Понятие временного ряда. Анализ временных рядов и его разделы. Цели, этапы и методы анализа временных рядов. (1,2)	1
	2	Детерминированная и случайная составляющая временного ряда. Тренд, сезонная и циклическая компонента. Модели тренда. (1,2)	2
	3	Метод скользящего среднего. (1,2,3)	2
	4	Метод экспоненциального сглаживания. (1,2,3)	2
Итого по разделу часов:			7
Раздел 5	1	Понятие, назначение и виды дисперсионного анализа. (1,2)	1
	2	Однофакторный дисперсионный анализ для связанных выборок. Однофакторный дисперсионный анализ для несвязанных выборок. (1,2,3)	3
	3	Факторный анализ. (1,2,3)	2
Итого по разделу часов:			6
Итого:			54

Примечание:

- 1 – проработка учебного материала (по конспектам лекций, учебной и научной литературе);
2 – подготовка к лабораторным занятиям, к тестированию;
3 – выполнение индивидуальных заданий.

5. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ)

По данной дисциплине курсовые проекты не предусмотрены.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Обеспечение обучающихся учебниками, учебными пособиями

№ п/п	Наименование учебника, учебного пособия	Автор	Год издания	Кол-во экз.	Электронная версия	Место размещения электронной версии
Основная литература						
1	Методы анализа данных: учеб. пособие	Макаров Р. И., Хорошева Е. Р.	2021	–	+	https://dspace.www1.vlsu.ru/handle/123456789/9369
2	Статистическая обработка результатов измерений: учебное пособие	Цыбрий И. К., Мороз К. А., Игнатенко В. И., Вяликов И. Л.	2021	–	+	https://e.lanbook.com/book/237854
3	Обработка экспериментальных данных: учебное пособие: в 2 частях. – Часть 1	Тазиева Р.Ф., Титов А. Н.	2017	–	+	https://e.lanbook.com/book/138538
4	Обработка экспериментальных данных: учебное пособие: в 2 частях. – Часть 2	Тазиева Р.Ф., Титов А. Н.	2018	–	+	https://e.lanbook.com/book/166193
5	Методы математической обработки экспериментальных данных учебно-методическое пособие	Гребенникова И.В.	2015	–	+	https://elar.ufrj.br/handle/10995/34780
Дополнительная литература						
1	Основы математической статистики. Обработка статистических данных в среде MS Excel: учебно-методическое пособие	Бельчикова О. Г.	2021	–	+	https://e.lanbook.com/book/197210
2	Эконометрика: лабораторный практикум в Excel: учебное пособие	Еремеева Н.С., Лебедева Т.С.	2016	–	+	https://e.lanbook.com/book/98143
3	Анализ и статистическая обработка исследовательских данных: учебно-методическое пособие	Фендель Т.В.	2017	–	+	https://e.lanbook.com/book/152744
4	Визуализация данных в Python. Работа с	Титов А.Н., Тазиева Р.	2022	–	+	https://e.lanbook.com/book/

	библиотекой Matplotlib: учебно-методическое пособие					331025
Итого по дисциплине: % печатных изданий – 0%; % электронных – 100%						

6.2. Программное обеспечение и Интернет-ресурсы

1. Google Colab
URL: <https://colab.research.google.com>
2. Видеоуроки по статистике
URL: <https://www.youtube.com/@irinamalykhina2590/videos>
3. Видеокурс по анализу данных
URL: https://www.youtube.com/playlist?list=PLDxLWC5qTf9Y-7r3w3h-Y0BgN6LP_tgUz

6.3. Методические указания и материалы по видам занятий

1. Никульчев Е. В., Алексеенко А. С., Ильин Д. Ю. Системы сбора и предобработки данных. Методы статистического анализа с использованием Google Colab: учебное пособие / Е.В. Никульчев, А.С. Алексеенко, Д.Ю. Ильин. – Москва: РТУ МИРЭА, 2023. (URL: <https://e.lanbook.com/book/382739>)

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Реализация данной учебной дисциплины осуществляется с использованием материально-технической базы, обеспечивающей проведение всех видов учебных занятий, предусмотренных программой учебной дисциплины и соответствующей действующим санитарным и противопожарным правилам и нормам.

Компьютерные классы оснащены современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест достаточно, чтобы обеспечивалась индивидуальная работа студента на отдельном персональном компьютере.

8. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

Данный курс предполагает овладение студентами теоретическими основами наиболее распространенных методов анализа данных, приобретение ими практических навыков применения программных средств, применяемых для анализа данных. Для успешного изучения материала необходимы знания таких основополагающих понятий как: виды случайных величин, характеристики случайных величин, основные законы распределения случайных величин, способы оценки выборочных характеристик, основы статистического вывода с использованием статистических критериев и проверки гипотез. Полученные в рамках курса знания являются основой для принятия эффективных управленческих решений в профессиональной деятельности.

Различные виды учебных занятий: лекции и лабораторные занятия – тесно связаны друг с другом. Материал, законспектированный на лекциях, необходимо регулярно дополнять сведениями из литературных и интернет-источников, представленных в рабочей программе дисциплины. Подготовка к лабораторным занятиям заключается в предварительном изучении соответствующего материала по конспекту лекций или по рекомендованной литературе.

Пропуск занятий, невыполнение лабораторных работ или неусвоение материала требуют компенсации путем самостоятельной работы студента. В случае необходимости следует обращаться к преподавателю за консультацией.

Планирование времени на самостоятельную работу, необходимого на изучение настоящей дисциплины, студентам лучше всего осуществлять на весь семестр, предусматривая при этом регулярное повторение пройденного материала.

9. ТЕХНОЛОГИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ

Курс 3 группа ФМ22ДР62ПФ1 (303) семестр 6

Преподаватель-лектор: *ст. преподаватель Калинкова Е.В.*

Преподаватель, ведущий лабораторные занятия: *ст. преподаватель Калинкова Е.В.*

Кафедра *Высшей и прикладной математики и информатики*

Семестр	Количество часов						Форма контроля
	Трудо- емкость, з.е./часы	В том числе					
		Аудиторных				СР	
		Всего	Л	ПЗ	ЛЗ		
6	3/108	54	18		36	54	Зачет с оценкой

Форма текущей аттестации	Расшифровка	Минималь- ное кол-во баллов	Максималь- ное ко-во баллов
Посещение лекционных занятий		0	10
Выполнение и защита лабораторных работ	За каждую лабораторную работу от 3 до 4 баллов	0	45
Тестирование		0	15
Итого количество баллов по текущей аттестации:		45	70
Промежуточная аттестация	Зачет	10	30
Итого по дисциплине:		55	100