

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ЛГПУ»)**

Структурное подразделение Институт физико-математического
образования, информационных и обслуживающих технологий
Кафедра информационных образовательных технологий и систем

УТВЕРЖДАЮ

Директор ИФМОИОТ

Е.Е. Горбенко

2023 г.



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Анализ больших данных

По направлению подготовки 09.04.04 Программная инженерия

Профиль подготовки Программное обеспечение систем и комплексов

Квалификация выпускника магистр

Форма обучения очная, заочная

Курс ОФО – 2 курс, ЗФО – 3 курс

Луганск, 2023

Рабочая программа учебной дисциплины является частью основной профессиональной образовательной программы для подготовки магистров по направлению подготовки 09.04.04 Программная инженерия очной и заочной форм обучения.

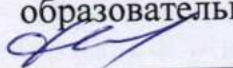
Рабочая программа учебной дисциплины разработана в соответствии с ФГОС ВО – магистратура по направлению подготовки 09.04.04 «Программная инженерия», утвержденным приказом Министерства образования и науки Российской Федерации от 19.09.2017 № 932 (с изменениями и дополнениями) и Профессиональным стандартом, утвержденным Приказом Министерства труда и социальной защиты Российской Федерации «Об утверждении профессионального стандарта 06.017 «Руководитель разработки программного обеспечения» от 20.07.2022 № 423н.

СОСТАВИТЕЛЬ:

доцент кафедры информационных образовательных технологий и систем,
кандидат физико-математических наук, доцент Швыров Вячеслав Владимирович

Утверждена на заседании кафедры информационных образовательных технологий и систем
Протокол от «24» ноября 2023 г. №8

Заведующий кафедрой информационных образовательных технологий и систем

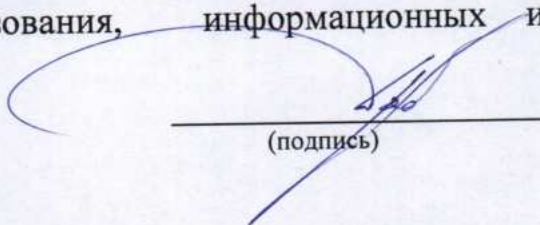

(подпись)

Д.А. Капустин

Одобрена на заседании учебно-методической комиссии Института физико-математического образования, информационных и обслуживающих технологий

Протокол от «06» декабря 2023 г. №5


Председатель учебно-методической комиссии Института физико-математического образования, информационных и обслуживающих технологий


(подпись)

О.В. Давыскиба

СОГЛАСОВАНО:

Заведующий учебно-методическим отделом


(подпись)

В.В. Савенков

1. Цели и задачи дисциплины

Цели изучения дисциплины: формирование компетенций в области использования информации, обработки и анализа ее автоматизированного анализа и визуализации.

Задачи:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- приобретение практических навыков работы большими данными.

2. Место дисциплины в структуре ОПОП

Учебная дисциплина «Анализ больших данных» относится к части, формируемой участниками образовательных отношений учебного плана (Б1.В.ДВ.04.01). Дисциплина реализуется кафедрой информационных образовательных технологий и систем (4) Институт физико-математического образования, информационных и обслуживающих технологий ФГБОУ ВО «ЛГПУ».

Необходимым условием для освоения учебной дисциплины являются знания фундаментальные понятия структуру классов Python; теоретические основы проектирования на Python; основные инструментальные для работы с Python; основные области применения приложений на Python; умения пользоваться своими знаниями для решения фундаментальных, прикладных и технологических задач инженерии знаний; делать правильные выводы из сопоставления результатов теории и практики; видеть в технических задачах математическое содержание; осваивать новые предметные области, теоретические подходы и практические методики; работать на современном компьютерном оборудовании и с новыми программными системами; эффективно использовать информационные технологии и компьютерную технику для достижения практически значимых результатов. навыки навыками освоения больших объемов информации, представленной в традиционной и электронной форме; навыками грамотной обработки результатов компьютерного моделирования и сопоставления их с теоретическими данными; практикой исследования и решения теоретических и прикладных задач; навыками теоретического анализа реальных задач, связанных с представлением и обработкой знаний.

Содержание дисциплины «Анализ больших данных» является логическим продолжением содержания дисциплин «Программирование», «Компьютерная дискретная математика», «Алгоритмы и структуры данных». и основой для дальнейшего освоения дисциплин: «Исследование и проектирование систем искусственного интеллекта», написания магистерской диссертации.

3. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Код по ФГОС ВО	Индикатор достижения	Результаты обучения по дисциплине
Универсальные		
Общепрофессиональные		
Профессиональные		
ПК-3. Владение навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов	<p>ПК-3.1. Знать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.2. Уметь использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.3. Владеть навыками использования методов создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p>	<p>ПК-3.1. Знает методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.2. Умеет использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.3. Владеет навыками использования методов создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p>

4. Структура и содержание дисциплины

4.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем часов (4 зач. ед.)	
	Очная форма	Заочная форма
Общая учебная нагрузка (всего)	144	144
Обязательная аудиторная учебная нагрузка (всего часов), в том числе:		
Лекции	16	4
Семинарские занятия		
Практические занятия		
Лабораторные работы	32	12
Курсовая работа / курсовой проект		
Другие формы организации учебного процесса (контрольные работы, индивидуальные занятия, консультации и др.)	36	9
Самостоятельная работа студента (всего)	60	9
Форма аттестация	Экзамен	Экзамен

4.2. Содержание дисциплины

Тема 1. Введение в большие данные.

Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.

Тема 2. Жизненный цикл аналитики данных.

Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.

Тема 3. Высокопроизводительные вычисления.

Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.

Тема 4. Масштабирование и многоуровневое хранение данных.

Использование пакетов Python для манипуляций с данными. Словари. Списки. Работа с Pandas.

Тема 5. Визуализация данных и результатов анализа.

Визуализация многомерных данных. Алгоритмы UMAP. TSNE

Тема 6. Классификация задач анализа данных.

Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.

Тема 7. Сложные методы аналитики.

Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов.

Тема 8. Основы построения нейронных сетей для анализа данных.

Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура.

4.3. Лекции

№ п/п	Название темы	Объем часов	
		Очная форма	Заочная форма
3-4 семестр / 7-8 триместр			
1	Тема 1. Введение в большие данные.	2	2
2	Тема 2. Жизненный цикл аналитики данных.	2	2
3	Тема 3. Высокопроизводительные вычисления.	2	
4	Тема 4. Масштабирование и многоуровневое	2	

	хранение данных.		
5	Тема 5. Визуализация данных и результатов анализа.	2	
6	Тема 6. Классификация задач анализа данных.	2	
7	Тема 7. Сложные методы аналитики.	2	
8	Тема 8. Основы построения нейронных сетей для анализа данных.	2	
Итого:		16	4

4.4. Практические занятия

4.5. Лабораторные работы

№ п/п	Название темы	Объем часов	
		Очная форма	Заочная форма
3-4 семестр / 7-8 триместр			
1	Установка Anaconda. Создание окружений	4	2
2	Установка пакетов	4	2
3	Визуализация многомерных данных в Matplotlib	4	2
4	Работа с JSON	4	2
5	Хранение больших данных в таблицах	4	2
6	Работа с Pandas	4	2
7	Пакеты для обработки данных	2	
8	Понятие датасета (набора данных). Проблемы классификации и кластеризации	2	
9	Классификация текстовых данных	2	
10	Большие данные и нейросети	2	
Итого:		32	12

4.6. Самостоятельная работа студентов

№ п/п	Название раздела / темы	Вид самостоятельной работы	Объем часов	
			Очная форма	Заочная форма
3-4 семестр / 7-8 триместр				
1	Работа с пакетом Numpy	Конспект лекций	12	2
2	Работа с пакетом Pandas	Конспект лекций	12	2
3	Работа с пакетом Matplotlib	Конспект лекций	12	2
4	Работа с графикой в пакете Seaborn	Конспект лекций	12	2
5	Базы данных в Python	Конспект лекций	12	1
Итого:			60	9

4.7. Курсовые работы / проекты

5. Методическое обеспечение, образовательные технологии

Преподавание дисциплины ведется с применением следующих видов образовательных технологий.

Наряду с методикой традиционной лекционно-практической работы предусмотрено использование активных форм и методов учебной деятельности, в том числе: учебные дискуссии, беседы, мозговой штурм.

Методика проблемно-диалогического обучения применяется в процессе лекционной работы над учебным материалом в каждой из тем учебной дисциплины.

Методика обучения в сотрудничестве с применением командных, групповых видов работы используется в процессе организации лабораторных работ.

Методика исследовательской деятельности используется как основа для организации самостоятельной работы студентов в объеме учебных тем.

Применяются средства мультимедиа: презентации, видео, базы ЭОР.

Информационные технологии: использование электронных образовательных ресурсов (электронный конспект, размещенный во внутренней сети или т.п.) при подготовке к лекциям, лабораторным работам и самостоятельной работе.

Работа в команде, проектная деятельность: совместная работа студентов в группе при выполнении лабораторных работ.

6. Формы контроля освоения дисциплины

Текущая аттестация студентов производится в дискретные временные интервалы в следующих формах: выполнение лабораторных работ; защита лабораторных работ.

Промежуточный контроль по результатам освоения дисциплины проходит в форме зачета (включает в себя ответ на теоретические вопросы и выполнение тестового задания).

Система оценивания учебных достижений студентов, оценочные средства представлены в фонде оценочных средств к рабочей программе учебной дисциплины (в приложении).

7. Учебно-методическое и программно-информационное обеспечение дисциплины

А) основная литература:

1. Келлехер Д. Наука о данных: базовый курс : учебное пособие. науч. ред. З. Мамедьяров; пер. с англ. М. Белоголовского. – Москва : Альпина Паблишер, 2020. – 224 с.

2. Кравченко Ю.А., Кулиев Э.В., Заруба Д.В. Тенденции развития компьютерных технологий: учебное пособие. Таганрог: Издательство Южного федерального университета, 2017. – 107с.
3. Бхаргава А. Грокаем алгоритмы: иллюстрированное пособие для программистов и любопытствующих. - СПб.: Питер, 2017. - 288 с.
4. Гэддис Т. Начинаем программировать на Python / пер. с англ. 4-е изд. - СПб.: БХВ-Петербург, 2019. - 768 с.
5. Мюллер Дж. Python для чайников. - СПб. : Диалектика, 2019. - 416 с.

Б) дополнительная литература:

1. Лутц М. Изучаем Python, пер. с англ. 3-е изд. - СПб.: Символ Плюс, 2009. - 848 с.
2. Рафгарден Т. Совершенный алгоритм. Жадные алгоритмы и динамическое программирование. - СПб.: Питер, 2020. - 256 с.
3. Рейтц К., Шлюссер Т. Автостопом по Python. - СПб. : Питер, 2017. - 336 с.
4. Фёдоров Д. Ю. Программирование на языке высокого уровня Python: учебное пособие для прикладного бакалавриата. - М. : Издательство Юрайт, 2019. - 161 с.

В) Интернет-ресурсы:

8. Материально-техническое обеспечение дисциплины

Лекционные занятия: комплект электронных презентаций/слайдов, аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) и т.п.

Лабораторные работы: компьютерный класс, оснащенный мультимедийным проектором, интерактивной доской, сетевой инфраструктурой и организованным доступом в Интернет, пакеты ПО MS Word, MS Excel , Anaconda, редактор Spider, пакеты Matplotlib..

Прочее: рабочее место преподавателя, оснащенное компьютером с доступом в Интернет, рабочие места студентов, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде и т.п.

9. Лист дополнений и изменений

[illegible]