

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 28.05.2025 14:17:01
Уникальный программный ключ:
sa953a01204891083f939673078ef1a989aae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Факультет физико-математических и естественных наук**
(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

СТАТИСТИЧЕСКАЯ ОБРАБОТКА БОЛЬШИХ ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И АНАЛИЗ ДАННЫХ

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Статистическая обработка больших данных» входит в программу магистратуры «Искусственный интеллект и анализ данных» по направлению 09.04.03 «Прикладная информатика» и изучается в 1 семестре 1 курса. Дисциплину реализует Кафедра теории вероятностей и кибербезопасности. Дисциплина состоит из 8 разделов и 16 тем и направлена на изучение

Целью освоения дисциплины является развитие у слушателей базовых знаний в области математической статистики больших данных и расширение навыков применения этих знаний для построения математических моделей случайных явлений и процессов, происходящих в реальной жизни, формирование у студентов необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать современные проблемы прикладной математики и информатики, проблемы обработки и анализа информации, а также умение разрабатывать и анализировать концептуальные и теоретические модели при решении научных и прикладных задач в области информационных технологий. Одна из главных проблем современной обработки и анализа данных - рост объемов данных, поэтому вопросам обработки большого объема данных посвящена данная дисциплина. Главная задача курса - сформировать целостное представление о современных проблемах анализа и обработки больших данных, помочь овладеть опытом разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных с применением моделей Data Mining. В данном курсе рассматриваются наиболее значимые с практической точки зрения разделы многомерного статистического анализа. Изучение данного курса является важной частью профессионального образования будущего магистра, специализирующегося в области прикладной информатики.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Статистическая обработка больших данных» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Знает принципы сбора, отбора и обобщения информации; УК-1.2 Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности; УК-1.3 Имеет практический опыт работы с информационными источниками, опыт научного поиска, создания научных текстов;
УК-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для	УК-7.1 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	
ОПК-1	Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, информатики и теории коммуникаций; ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические и информационные объекты; ОПК-1.3 Имеет практический опыт работы с решением математических и информационных задач и применяет его в профессиональной деятельности;
ОПК-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1 Знает основные положения и концепции в области программирования, языков программирования, теории коммуникации, знает основную терминологию, знаком с перечнем ПО, включенного в Единый Реестр Российских программ; ОПК-2.2 Умеет анализировать типовые языки программирования, составлять программы; ОПК-2.3 Имеет практический опыт решения задач анализа и интеграции различных типов программного обеспечения, анализа типов коммуникации;
ПК-2	Организационное и технологическое обеспечение проектирования и дизайна ИС	ПК-2.3 Знает основы программирования, современные методики тестирования разрабатываемых информационных систем, современные инструменты и методы верификации программного кода, теорию баз данных, системы хранения и анализа данных, инструменты и методы проектирования баз данных; ПК-2.4 Знает методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов, основы реинжиниринга бизнес-процессов организации, программные средства и платформы инфраструктуры информационных технологий организации, современные подходы и стандарты автоматизации организаций;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Статистическая обработка больших данных» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Статистическая обработка больших данных».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-7	Способен: искать нужные источники информации и данные, воспринимать,		Языки программирования для задач математической физики;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных		Интеллектуальный анализ больших данных; Интеллектуальные системы и их применение; Глубокое обучение в обработке изображений; Обучение с подкреплением; Practicum in Artificial Intelligence; Преддипломная практика; Научно-исследовательская работа;
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий		Технологическая (проектно-технологическая) практика; Преддипломная практика; Научно-исследовательская работа; Прикладные методы компьютерной лингвистики; Анализ мультимодальных данных; Интеллектуальный анализ больших данных; Методы интеллектуального анализа текстов; Распознавание образов и обработка изображений; Интеллектуальные системы и их применение; Глубокое обучение в обработке изображений; Языки программирования для задач математической физики; Линейное программирование и оптимизация; Вычислительные методы; Обучение с подкреплением; Practicum in Artificial Intelligence;
ОПК-1	Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном		Технологическая (проектно-технологическая) практика; Ознакомительная практика; Преддипломная практика; Научно-исследовательская работа; Методы интеллектуального анализа текстов; Языки программирования для задач математической физики; Вычислительные методы;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	контексте		Пакеты символьных вычислений; Прикладные методы компьютерной лингвистики;
ОПК-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач		Программирование приложений для анализа данных; Линейное программирование и оптимизация; Языки программирования для задач математической физики; Технологическая (проектно-технологическая) практика; Ознакомительная практика; Преддипломная практика; Научно-исследовательская работа;
ПК-2	Организационное и технологическое обеспечение проектирования и дизайна ИС		Технологическая (проектно-технологическая) практика; Ознакомительная практика; Анализ мультимодальных данных; Интеллектуальный анализ больших данных; Обучение с подкреплением; Программирование приложений для анализа данных; Линейное программирование и оптимизация; Прикладные методы компьютерной лингвистики; Интеллектуальные системы и их применение; Языки программирования для задач математической физики; Методы интеллектуального анализа текстов; Распознавание образов и обработка изображений; Глубокое обучение в обработке изображений;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Статистическая обработка больших данных» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			1
<i>Контактная работа, ак.ч.</i>	36		36
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	18		18
Практические/семинарские занятия (СЗ)	0		0
<i>Самостоятельная работа обучающихся, ак.ч.</i>	81		81
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.	1.1	Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.	ЛК, ЛР
		1.2	Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации.	ЛК, ЛР
		1.3	Программное обеспечение в области анализа данных. Аналитические платформы.	ЛК, ЛР
Раздел 2	Дисперсионный анализ	2.1	Основная задача дисперсионного анализа (ДА). Однофакторный ДА. Формула разложения выборочной дисперсии. Леммы о несмещенных оценках генеральной дисперсии. Проверка основной гипотезы ДА. Выборочный коэффициент детерминации.	ЛК, ЛР
Раздел 3	Корреляционно-регрессионный анализ	3.1	Понятие функциональной, стохастической и корреляционной зависимости. Функция регрессии. Генеральное корреляционное отношение. Его свойства. Выборочное корреляционное отношение. Проверка гипотезы о его значимости.	ЛК, ЛР
		3.2	Линейная парная регрессия. Уравнение для выборочной линии регрессии. Выборочный коэффициент корреляции. Классическая парная нормальная линейная регрессионная модель. Оценка значимости уравнения регрессии. Классическая модель множественной регрессии.	ЛК, ЛР
		3.3	Метод наименьших квадратов для модели множественной регрессии. Оценка дисперсии погрешностей в модели множественной регрессии. Оптимальный выбор матрицы плана.	ЛК, ЛР
		3.4	Интервальные оценки параметров нормальной классической модели множественной регрессии. Оценка значимости уравнения множественной регрессии. Коэффициент детерминации.	ЛК, ЛР
		3.5	Особенности практического применения регрессионных моделей. Мультиколлинеарность. Нелинейные регрессионные модели. Обобщенная линейная модель множественной регрессии.	ЛК, ЛР
		3.6	Прогнозирование с помощью линейной регрессии	ЛК, ЛР
Раздел 4	Кластерный анализ	4.1	Расстояния между объектами и кластерами. Понятие кластерного анализа.	ЛК, ЛР
		4.2	Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.	ЛК, ЛР
Раздел 5	Факторный анализ	5.1	Основные компоненты факторного анализа.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
			Методика факторного анализа в случае одного и нескольких факторов.	
Раздел 6	Дискриминантный анализ	6.1	Методы дискриминантного анализа. Критерии сравнения выборок по нескольким признакам	ЛК, ЛР
Раздел 7	Нейронные сети	7.1	Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура.	ЛК, ЛР
Раздел 8	Деревья решений	8.1	Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура.	ЛК, ЛР

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	Компьютер/ноутбук с доступом сети Интернет и электронно-образовательной среде Университета, браузер, ПО для просмотра PDF, Яндекс Телемост.
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 15 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	ОС Linux/Windows, VirtualBox, GNS3, Wireshark, Octave. Дополнительное ПО: офисный пакет MS Office или LibreOffice, OBS Studio
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	ОС Linux/Windows, VirtualBox, GNS3, Wireshark, Octave. Дополнительное ПО: офисный пакет MS Office или LibreOffice, OBS Studio

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Макшанов А.В., Журавлев А.Е., Тындыкарь Л.Н. Большие данные. Big data: учебник для вузов. 2021. Изд-во Лань. 186 с.

2. Ватьян А.С., Добренко Н.В., Гусарова Н.Ф. Data science: проблемы и решения. Изд-во Национального Исследовательский Университета ИТМО, 2025. 221 с.
3. Путко, Б.А. Эконометрика : учебник / Б.А. Путко, Н.Ш. Кремер ; ред. Н.Ш. Кремер. - 3-е изд., перераб. и доп. - Москва : Юнити-Дана, 2012. - 329 с. - (Золотой фонд российских учебников). - ISBN 978-5-238-01720-4 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=118251>
4. Бочаров Павел Петрович. Теория вероятностей и математическая статистика [текст] : Учебное пособие / П.П. Бочаров, А.В. Печинкин. - М. : Физматлит, 2005. - 295 с. : ил. - ISBN 5-9221-0633-3 : 153.00. (ЕТ 100)
5. Буравлев Александр Иванович. Эконометрика [Текст/электронный ресурс] : Учебное пособие / А.И. Буравлев. - Электронные текстовые данные. - М. : БИНОМ. Лаборатория знаний, 2019. - 164 с. : ил. - ISBN 978-5-9963-0741-8 : 220.00. Режим доступа: <http://lib.rudn.ru/ProtectedView/Book/ViewBook/6768>

Дополнительная литература:

1. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие / В.Е. Гмурман. - Изд. 3-е, перераб. и доп. - Москва : Высшая школа, 1979. - 400 с. : ил. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=458330>
2. Гмурман Владимир Ефимович. Руководство к решению задач по теории вероятностей и математической статистике [Текст] : Учебное пособие для вузов / В.Е. Гмурман. - 4-е изд., стереотип. - М. : Высшая школа, 1998. - 400 с. : ил. - ISBN 5-06-003465-8 : 13.00. (ЕТ 44)

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров
 - Электронно-библиотечная система РУДН – ЭБС РУДН <https://mega.rudn.ru/MegaPro/Web>
 - ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
 - ЭБС «Юрайт» <http://www.biblio-online.ru>
 - ЭБС «Консультант студента» www.studentlibrary.ru
 - ЭБС «Знаниум» <https://znanium.ru/>
2. Базы данных и поисковые системы
 - Sage <https://journals.sagepub.com/>
 - Springer Nature Link <https://link.springer.com/>
 - Wiley Journal Database <https://onlinelibrary.wiley.com/>
 - Наукометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Статистическая обработка больших данных».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИК:

Доцент кафедры теории
вероятностей и
кибербезопасности

Должность, БУП

Подпись

Матюшенко Сергей
Иванович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой
математического
моделирования и
искусственного интеллекта

Должность БУП

Подпись

Малых Михаил
Дмитриевич

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Заведующий кафедрой
математического
моделирования и
искусственного интеллекта

Должность, БУП

Подпись

Малых Михаил
Дмитриевич

Фамилия И.О.