

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 29.05.2023 12:43:55

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a9896ae18a

Федеральное государственное автономное образовательное учреждение высшего образования

«Российский университет дружбы народов»

Факультет физико-математических и естественных наук

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Обработка больших данных с использованием машинного обучения

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки:

09.03.03 Прикладная информатика

(код и наименование направления подготовки/специальности)

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

Прикладная информатика

(наименование (профиль/специализация) ОП ВО)

2023 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Обработка больших данных с использованием машинного обучения» является введение учащихся в предметную область современных компьютерных технологий и компьютерного моделирования на основе анализа больших данных в машинном обучении. В рамках освоения данной дисциплины студенты ознакомятся с линейными моделями анализа больших данных, байесовскими методами, искусственными нейронными сетями и алгоритмами кластеризации.

В результате обучения студенты приобретут умения и навыки правильно оценить сложность научно-исследовательских заданий при разработке прикладных компьютерных моделей, связанных с анализом больших данных, аргументировано выбирать методы решения поставленных задач, а затем эффективно выполнять компьютерное моделирование на основе машинного обучения, а также все необходимые вычисления в рамках поставленных прикладных задач.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Обработка больших данных с использованием машинного обучения» направлено на формирование у обучающихся следующих компетенций (части компетенций): УК-1; ОПК-1; ОПК-10; ПК-7

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	Способность осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Знает принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач.
		УК-1.2. Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности.
		УК-1.3. Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений.
ОПК-1	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального	ОПК-1.1. Знает основы математики, физики, вычислительной техники и программирования.
		ОПК-1.2. Умеет решать стандартные профессиональные задачи с применением естественнонаучных и обще-инженерных знаний, методов математического анализа и моделирования
		ОПК-1.3. Владеет навыками теоретического и экспериментального исследования объектов

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	исследования в профессиональной деятельности	профессиональной деятельности
ОПК-10	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	<p>ОПК-10.1 Знает базовые принципы цифровых технологий и методов, необходимых в профессиональной деятельности в области прикладной информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.</p> <p>ОПК-10.2 Умеет применять необходимые в профессиональной деятельности цифровые технологии и методы в области прикладной информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.</p> <p>ОПК-10.3 Владеет необходимыми в профессиональной деятельности технологиями и методами в области прикладной информатики для: изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации и пр.</p>
ПК-7	Проведение работ по обработке и анализу научно-технической информации и результатов исследований	<p>ПК-7.1. Знает основы научно-исследовательской деятельности в области информационных технологий, знает основные методы решения прикладных задач, современные методы информационных технологий</p> <p>ПК-7.2. Умеет применять полученные знания для решения стандартных задач в области информационных технологий и в собственной научно-исследовательской деятельности</p> <p>ПК-7.3. Имеет практический опыт научно-исследовательской деятельности в области информационных технологий</p> <p>ПК-7.4. Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации; владеет базовыми навыками подготовки научных обзоров и (или) публикаций, рефератов и библиографий по тематике проводимых исследований на русском и иностранном языке</p>

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Обработка больших данных с использованием машинного обучения» относится к части, формируемой участниками образовательных отношений блока Б1. ОП ВО.

В рамках ОП ВО обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Обработка больших данных с использованием машинного обучения».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики ¹
УК-1	Способность осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	Вычислительные системы, сети и телекоммуникации Основы информационной безопасности История Философия Социальные и этические вопросы информационных технологий Интеллектуальные системы Теоретические основы информатики Имитационное моделирование Алгоритмы машинной графики и обработки изображений Управление проектами разработки информационных систем Компьютерный практикум по моделированию Методы машинного обучения для анализа временных рядов и панельных данных	Технологии интеллектуального анализа данных и прогнозирование Кибербезопасность предприятия Основы проектирования сетей и систем телекоммуникаций Компьютерный практикум по статистическому анализу данных Компьютерный практикум по интеллектуальным системам Анализ больших данных при моделировании сложно-структурированных систем Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы) Технологическая (проектно-технологическая) практика Научно-исследовательская работа Преддипломная практика

¹ - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики
ОПК-1	Способность применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;	<p>Математический анализ Линейная алгебра Дискретная математика и математическая логика Дифференциальные и разностные уравнения Теория вероятностей и математическая статистика Теория конечных графов Технология программирования Python и его приложения Вычислительные системы, сети и телекоммуникации Концепции современного естествознания Интеллектуальные системы Теоретические основы информатики Структуры данных и парадигмы программирования Теория автоматов и формальных языков Алгоритмы машинной графики и обработки изображений Линейный и нелинейный регрессионный анализ больших данных Методы машинного обучения для анализа временных рядов и панельных данных</p>	<p>Технологии интеллектуального анализа данных и прогнозирование Кибербезопасность предприятия Основы проектирования сетей и систем телекоммуникаций Компьютерный практикум по моделированию Компьютерный практикум по информационным технологиям Параллельное программирование Прикладной анализ данных с использованием языка Python Компьютерный практикум по статистическому анализу данных Компьютерный практикум по интеллектуальным системам Анализ больших данных при моделировании сложно-структурированных систем Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы) Технологическая (проектно-технологическая) практика Научно-исследовательская работа Преддипломная практика</p>
ОПК-10	Способность понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной	<p>Основы программирования Технология программирования Python и его приложения Вычислительные системы, сети и телекоммуникации</p>	<p>Системы управления базами данных Технологии интеллектуального анализа данных и прогнозирование Кибербезопасность предприятия Основы проектирования</p>

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики
	деятельности	Архитектура компьютеров и операционные системы Интеллектуальные системы Теоретические основы информатики Основы Web-технологий Основы администрирования операционных систем Логическое программирование Структуры данных и парадигмы программирования Реляционные базы данных Теория автоматов и формальных языков Управление проектами разработки информационных систем Сетевые технологии Администрирование сетевых подсистем Компьютерный практикум по моделированию Компьютерный практикум по информационным технологиям Методы машинного обучения для анализа временных рядов и панельных данных	сетей и систем телекоммуникаций Параллельное программирование Прикладной анализ данных с использованием языка Python Компьютерный практикум по статистическому анализу данных Компьютерный практикум по интеллектуальным системам Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы) Технологическая (проектно-технологическая) практика Научно-исследовательская работа Преддипломная практика
ПК-7	Проведение работ по обработке и анализу научно-технической информации и результатов исследований	Линейный и нелинейный регрессионный анализ больших данных Методы машинного обучения для анализа временных рядов и панельных данных	Технологии интеллектуального анализа данных и прогнозирование Основы проектирования сетей и систем телекоммуникаций Компьютерный практикум по статистическому анализу данных

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики
			Компьютерный практикум по интеллектуальным системам Анализ больших данных при моделировании сложно-структурированных систем Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы); Технологическая (проектно-технологическая) практика Научно-исследовательская работа Преддипломная практика

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Обработка больших данных с использованием машинного обучения» составляет 4 зачетных единиц.

Таблица 4.1. Виды учебной работы по периодам освоения ОП ВО

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)
		6
<i>Контактная работа, ак.ч.</i>	54	54
в том числе:		
Лекции (ЛК)	18	18
Лабораторные работы (ЛР)	-	-
Практические/семинарские занятия (СЗ)	36	36
<i>Самостоятельная работа обучающихся, ак.ч.</i>	63	63
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27	27
Общая трудоемкость дисциплины	ак.ч.	144
	зач.ед.	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

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

Наименование раздела дисциплины	Содержание раздела (темы)	Вид учебной работы ²
Раздел 1. Основные	Тема 1.1. Постановка задач обучения. Объекты и	ЛК, СЗ

² - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – семинарские занятия.

Наименование раздела дисциплины	Содержание раздела (темы)	Вид учебной работы
понятия и определения.	признаки, вектора признаков, классы.	
	Тема 1.2. Типы задач: классификация, регрессия, кластеризация, ранжирование.	ЛК, СЗ
	Тема 1.3. Основные понятия: методы обучения, функция потерь и функционал качества, переобучение, обобщающая способность. Примеры прикладных задач.	ЛК, СЗ
Раздел 2. Линейные модели.	Тема 2.1. Метод наименьших квадратов (МНК). Аналитическое и численное решение задачи МНК. Градиентный спуск, методы оценивания градиента. Функции потерь.	ЛК, СЗ
	Тема 2.2. Регуляризация. Методы оценивания обобщающей способности, кроссвалидация. Метрики качества регрессии. Линейная классификация.	ЛК, СЗ
	Тема 2.3. Случай двух классов и случай произвольного числа классов. Обучение линейного классификатора. Метод опорных векторов. Логистическая регрессия.	ЛК, СЗ
Раздел 3. Решающие деревья и композиции алгоритмов.	Тема 3.1. Общий алгоритм построения, критерии информативности. Конкретные критерии для классификации и регрессии.	ЛК, СЗ
	Тема 3.2. Тонкости решающих деревьев: обработка пропущенных значений, стрижка, регуляризация. Разложение ошибки на смещение и разброс.	ЛК, СЗ
	Тема 3.3. Бэггинг и бустинг. Случайные леса. Градиентный бустинг над решающими деревьями.	ЛК, СЗ
Раздел 4. Байесовские методы.	Тема 4.1. Наивный байесовский классификатор, его связь с линейными моделями.	ЛК, СЗ
	Тема 4.2. Параметрическое оценивание плотности.	ЛК, СЗ
	Тема 4.3. Непараметрическое оценивание плотности.	ЛК, СЗ
Раздел 5. Искусственные нейронные сети.	Тема 5.1. Проблема полноты. Задача исключающего "или". Вычислительные возможности двух и трехслойных сетей.	ЛК, СЗ
	Тема 5.2. Метод обратного распространения ошибки. Машина Больцмана, машина Гемгольца и алгоритмы их обучения. Нейронные сети Кохонена. Сеть радиально базисных функций. Алгоритмы глубокого обучения.	ЛК, СЗ
	Тема 5.3. Алгоритмы кластеризации. Задача кластеризации. Эвристические графовые алгоритмы. Метод k-средних.	ЛК, СЗ
Раздел 6. Обработка	Тема 6.1. Процесс анализа больших данных.	ЛК, СЗ

Наименование раздела дисциплины	Содержание раздела (темы)	Вид учебной работы
больших данных. Основы работы с большими массивами информации.	Технологии анализа больших данных. Научные проблемы в области больших данных.	
	Тема 6.2. Методы и техники анализа, применимые к большим данным. Аналитические методы, применимые к большим объемам данных. Многомерные связи, ассоциации, корреляции. Обработка неструктурированных данных.	ЛК, СЗ
	Тема 6.3. Технические сложности работы с большими данными и методы их решения.	ЛК, СЗ

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, MS Teams. Дополнительное ПО: офисный пакет MS Office или LibreOffice.
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве числа обучающихся), доской (экраном) и техническими средствами мультимедиа презентаций.	ОС Windows или Linux Python 3, GNU Octave, Scilab. Дополнительное ПО: офисный пакет MS Office или LibreOffice
Для самостоятельной работы обучающихся	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	Компьютер/ноутбук с доступом сети Интернет и электронно-образовательной среде Университета, браузер, ПО для просмотра PDF, MS Teams, Python 3, GNU Octave, Scilab.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Мхитарян В. С. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>.
2. Миркин Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469306>.
3. Платонов А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2022. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/508804>.
4. Сидняев Н. И. Теория планирования эксперимента и анализ статистических данных : учебник и практикум для вузов / Н. И. Сидняев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 495 с. — (Высшее образование). — ISBN 978-5-534-05070-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/508082>.

Дополнительная литература:

1. Воронов В.И. Data Mining - технологии обработки больших данных : учебное пособие / Воронов В.И., Воронова Л.И., Усачев В.А.. — Москва : Московский технический университет связи и информатики, 2018. — 47 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/81324.html>.
2. Гателюк О. В. Численные методы : учеб. пособие для академического бакалавриата / О. В. Гателюк, Ш. К. Исмаилов, Н. В. Манюкова. — Москва : Издательство Юрайт, 2019. — 140 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-534-05894-9.
3. Железнов М.М. Методы и технологии обработки больших данных : учебно-методическое пособие / Железнов М.М.. — Москва : МИСИ-МГСУ, ЭБС АСВ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/101802.html>.
4. Кузнецов В. В. Системный анализ: учебник и практикум для вузов / В. В. Кузнецов [и др.] ; под общей редакцией В. В. Кузнецова. — Москва : Издательство Юрайт, 2022. — 270 с. — (Высшее образование). — ISBN 978-5-9916-8591-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490660>.
5. Лаврищева Е. М. Программная инженерия и технологии программирования сложных систем : учебник для вузов / Е. М. Лаврищева. — 2-е изд., испр. и доп.

- Москва : Издательство Юрайт, 2019. — 432 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-07604-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/436514>.
6. Третьяк, Л. Н. Основы теории и практики обработки экспериментальных данных : учебное пособие для вузов / Л. Н. Третьяк, А. Л. Воробьев ; под общей редакцией Л. Н. Третьяк. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 237 с. — (Высшее образование). — ISBN 978-5-534-08623-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/492913>
7. Пяткина Д.А. Математическое моделирование в экономике и финансах : учебно- методическое пособие для студентов, обучающихся по специальности "Бизнес-информатика" / Д.А. Пяткина, С.И. Матюшенко. - Электронные текстовые данные. - М. : РУДН, 2018. - 40 с. : ил. - ISBN 978-5-209-08322-1 : 71.04.

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 <http://lib.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 «Лань» <http://e.lanbook.com/>
- ЭБС РГБ <http://www.rsl.ru/>

2. Базы данных и поисковые системы:

- электронный фонд правовой и нормативно-технической документации <http://docs.cntd.ru/>

- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS <http://www.elsevierscience.ru/products/scopus/>
- реферативная база данных Science Direct <http://www.sciencedirect.com> Описание: Ресурс содержит коллекцию научной, технической полнотекстовой и библиографической информации. База данных мультидисциплинарного характера включает научные журналы по точным и техническим наукам.
- реферативная база данных EBSCO <http://search.ebscohost.com>, Academic Search Premier (база данных комплексной тематики, содержит информацию по гуманитарным и естественным областям знания).
- реферативная база данных Springer/Kluwer <http://www.springerlink.com>. Журналы и книги издательства Springer/Kluwer охватывают различные области знания и разбиты на предметные категории.
- реферативная база данных Tailor & Francis <http://www.informaworld.com> . Коллекция журналов насчитывает более 1000 именованных по всем областям знаний.
- реферативная база данных Zentralblatt MATH (zbMATH) <https://zbmath.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля³:




1. Курс лекций по дисциплине «Обработка больших данных с использованием машинного обучения».

2. Сборник задач по дисциплине «Обработка больших данных с использованием машинного обучения».

8. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ И БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ ПО ДИСЦИПЛИНЕ

Оценочные материалы и балльно-рейтинговая система⁴ оценивания уровня сформированности компетенций (части компетенций) по итогам освоения дисциплины «Обработка больших данных с использованием машинного обучения» представлены в Приложении к настоящей Рабочей программе дисциплины.

РАЗРАБОТЧИКИ:

Доцент кафедры прикладной информатики и теории вероятностей		С.А.Васильев
Должность, БУП	Подпись	Фамилия И.О.
РУКОВОДИТЕЛЬ БУП: Зав. кафедрой прикладной информатики и теории вероятностей		К.Е. Самуйлов
Наименование БУП	Подпись	Фамилия И.О.
РУКОВОДИТЕЛЬ ОП ВО: Зав. кафедрой информационных технологий		Ю.Н. Орлов
Наименование БУП	Подпись	Фамилия И.О.

³ - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины в ТУИС

⁴ - ОМ и БРС формируются на основании требований соответствующего локального нормативного акта РУДН.