

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 22.05.2024 14:54:40

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a989dae18a

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

ОБРАБОТКА БОЛЬШИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(код и наименование направления подготовки/специальности)

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

ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(наименование (профиль/специализация) ОП ВО)

2024 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Обработка больших данных с использованием машинного обучения» входит в программу бакалавриата «Фундаментальная информатика и информационные технологии» по направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 6 семестре 3 курса. Дисциплину реализует Кафедра математического моделирования и искусственного интеллекта. Дисциплина состоит из 6 разделов и 18 тем и направлена на изучение предметной области современных компьютерных технологий и компьютерного моделирования на основе анализа больших данных в машинном обучении.

Целью освоения дисциплины является 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):

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

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-12	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	УК-12.1 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;
ПК-1	Способен разрабатывать и отлаживать программный код	ПК-1.1 Знает основы программирования; современные объектно-ориентированные языки программирования; современные структурные языки программирования; языки современных бизнес-приложений; ПК-1.2 Умеет кодировать на языках программирования; тестировать результаты кодирования; ПК-1.3 Владеет навыками разработки кода информационной системы; навыками верификации кода информационной системы;
ПК-2	Способен выполнять работы и управлять работами по	ПК-2.1 Знает инструменты и методы разработки архитектуры, проектирования и дизайна информационных систем;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	инструменты верификации программного кода; ПК-2.2 Умеет проектировать и верифицировать архитектуру информационной системы; кодировать на языках программирования; тестировать результаты кодирования; ПК-2.3 Владеет инструментами и методами проектирования и верификации архитектуры информационной системы, разработки и верификации структуры программного кода информационной системы;
ПК-3	Способен осуществлять администрирование прикладного программного обеспечения, сетевой подсистемы и систем управления базами данных инфокоммуникационной системы организации	ПК-3.1 Знает основы архитектуры, устройства и функционирования информационно-вычислительных систем и сетевых подсистем инфокоммуникационной системы организации; методику установки и администрирования программных систем и сетевых подсистем инфокоммуникационной системы организации; ПК-3.2 Умеет настраивать и администрировать программные системы, сетевые подсистемы и базы данных инфокоммуникационной системы организации; ПК-3.3 Имеет практический опыт эксплуатации и администрирования программных систем, сетевых подсистем и баз данных инфокоммуникационной системы организации;
ПК-4	Способен проводить работы по обработке и анализу научно-технической информации и результатов исследований	ПК-4.1 Знает основы научно-исследовательской деятельности в области информационных технологий, основные методы решения прикладных задач, современные методы информационных технологий; принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации; ПК-4.2 Умеет применять полученные знания для решения стандартных задач в области информационных технологий и в собственной научно-исследовательской деятельности; ПК-4.3 Владеет базовыми навыками подготовки научных обзоров и (или) публикаций, рефератов и библиографий по тематике проводимых исследований на русском и иностранном языке;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Обработка больших данных с использованием машинного обучения» относится к блоку по выбору блока образовательной программы высшего образования.

В рамках образовательной программы высшего образования 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*
УК-12	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов	Обработка данных и визуализация; Основы машинного обучения и нейронные сети;	Технологическая (проектно-технологическая) практика; Научно-исследовательская работа; Преддипломная практика; Технологии интеллектуального анализа данных и прогнозирование;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	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		<i>Компьютерный практикум по интеллектуальным системам**;</i> <i>Компьютерный практикум по статистическому анализу данных**;</i> Анализ больших данных;
ПК-1	Способен разрабатывать и отлаживать программный код	Основы программирования; Технология программирования; Технологии искусственного интеллекта; Методы машинного обучения для анализа временных рядов и панельных данных; Методы искусственного интеллекта; Алгоритмы машинной графики и обработки изображений; Основы машинного обучения и нейронные сети; Теория автоматов и формальных языков; Программная инженерия; Архитектура компьютеров и операционные системы; Компьютерная алгебра; Обработка данных и визуализация; Компьютерная геометрия; Машинное обучение в телекоммуникациях;	Анализ больших данных; Технологии интеллектуального анализа данных и прогнозирование; Интеллектуальные системы; Моделирование сетей передачи данных; <i>Параллельное программирование**;</i> <i>Прикладной анализ данных с использованием языка Python**;</i> <i>Компьютерный практикум по интеллектуальным системам**;</i> <i>Компьютерный практикум по статистическому анализу данных**;</i> Анализ больших данных при моделировании сложно-структурированных систем; Научно-исследовательская работа; Технологическая (проектно-технологическая) практика; Преддипломная практика;
ПК-2	Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	Основы программирования; Технология программирования; Технологии искусственного интеллекта; Методы искусственного интеллекта; Алгоритмы машинной графики и обработки изображений; Программная инженерия; Машинное обучение в телекоммуникациях;	Технологическая (проектно-технологическая) практика; Преддипломная практика; Интеллектуальные системы; <i>Компьютерный практикум по интеллектуальным системам**;</i> Анализ больших данных при моделировании сложно-структурированных систем;
ПК-3	Способен осуществлять администрирование прикладного программного обеспечения, сетевой подсистемы и систем управления базами данных инфокоммуникационной	Сетевые технологии; Администрирование сетевых подсистем; Реляционные базы данных; Вычислительные системы, сети и телекоммуникации; Архитектура компьютеров и операционные системы;	Моделирование сетей передачи данных; Кибербезопасность предприятия; Системы управления базами данных; Технологическая (проектно-технологическая) практика;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	системы организации	Основы информационной безопасности; Машинное обучение в телекоммуникациях;	Преддипломная практика;
ПК-4	Способен проводить работы 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; Методы искусственного интеллекта; Машинное обучение в телекоммуникациях; Программная инженерия; Прикладные задачи теории массового обслуживания; <i>Практический курс профессионального перевода**;</i> <i>Практический курс профессионального перевода (русский язык как иностранный)**;</i>	Научно-исследовательская работа; Преддипломная практика; Интеллектуальные методы разделения сетевых ресурсов; Технологии интеллектуального анализа данных и прогнозирование; Интеллектуальные системы; Анализ больших данных при моделировании сложно-структурированных систем; Моделирование сетей передачи данных; <i>Компьютерный практикум по интеллектуальным системам**;</i> <i>Компьютерный практикум по статистическому анализу данных**;</i>

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Обработка больших данных с использованием машинного обучения» 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)
			6
<i>Контактная работа, ак.ч.</i>	54		54
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	36		36
<i>Самостоятельная работа обучающихся, ак.ч.</i>	54		54
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	36		36
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	Основные понятия и определения	1.1	Постановка задач обучения. Объекты и признаки, вектора признаков, классы.	ЛК, СЗ
		1.2	Типы задач: классификация, регрессия. кластеризация, ранжирование.	ЛК, СЗ
		1.3	Основные понятия: методы обучения, функция потерь и функционал качества, переобучение, обобщающая способность. Примеры прикладных задач.	ЛК, СЗ
Раздел 2	Линейные модели	2.1	Метод наименьших квадратов (МНК). Аналитическое и численное решение задачи МНК. Градиентный спуск, методы оценивания градиента. Функции потерь.	ЛК, СЗ
		2.2	Регуляризация. Методы оценивания обобщающей способности, кроссвалидация. Метрики качества регрессии. Линейная классификация.	ЛК, СЗ
		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	3.1	Общий алгоритм построения, критерии информативности. Конкретные критерии для классификации и регрессии.	ЛК, СЗ
		3.2	Тонкости решающих деревьев: обработка пропущенных значений, стрижка, регуляризация. Разложение ошибки на смещение и разброс.	ЛК, СЗ
		3.3	Бэггинг и бустинг. Случайные леса. Градиентный бустинг над решающими деревьями.	ЛК, СЗ
Раздел 4	Байесовские методы	4.1	Наивный байесовский классификатор, его связь с линейными моделями.	ЛК, СЗ
		4.2	Параметрическое оценивание плотности.	ЛК, СЗ
		4.3	Непараметрическое оценивание плотности.	ЛК, СЗ
Раздел 5	Искусственные нейронные сети	5.1	Проблема полноты. Задача исключающего "или". Вычислительные возможности двух и трехслойных сетей.	ЛК, СЗ
		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
		5.3	Алгоритмы кластеризации. Задача кластеризации. Эвристические графовые алгоритмы. Метод k-средних.	ЛК, СЗ
Раздел 6	Обработка больших данных. Основы работы с большими массивами информации	6.1	Процесс анализа больших данных. Технологии анализа больших данных. Научные проблемы в области больших данных.	ЛК, СЗ
		6.2	Методы M ^o техники анализа, применимые к большим данным. Аналитические методы, применимые к большим объемам данных. Многомерные связи, ассоциации, корреляции. Обработка неструктурированных данных.	ЛК, СЗ
		6.3	Технические сложности работы с большими данными и методы их решения.	ЛК, СЗ

* - 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, MS Teams. Дополнительное ПО: офисный пакет MS Office или LibreOffice
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели 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	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.

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

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: [snekupoHttbiii 1/ Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: http://www.iprbookshop.ru/101802.html](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 «Троицкий мост»

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.elsevier.com/locate/elsevier/scopus/>

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

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

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

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

Оценочные материалы и балльно-рейтинговая система* 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.

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

РАЗРАБОТЧИК:

Доцент кафедры
математического
моделирования и
искусственного интеллекта

Должность, БУП

Подпись

Васильев Сергей
Анатольевич

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 теории
вероятностей и
кибербезопасности

Должность, БУП

Подпись

Самуйлов Константин
Евгеньевич

Фамилия И.О.