

Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»

Направление подготовки/специальность: 01.04.02 Прикладная математика и информатика

Наименование образовательной программы: Искусственный интеллект

Уровень образования: высшее образование - магистратура

Форма обучения: Очная


Рабочая программа дисциплины
ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Блок:	Блок 1 «Дисциплины (модули)»
Часть образовательной программы:	Часть, формируемая участниками образовательных отношений
№ дисциплины по учебному плану:	Б1.Ч.06.04.01
Трудоемкость в зачетных единицах:	3 семестр - 4;
Часов (всего) по учебному плану:	144 часа
Лекции	3 семестр - 16 часов;
Практические занятия	3 семестр - 32 часа;
Лабораторные работы	не предусмотрено учебным планом
Консультации	3 семестр - 2 часа;
Самостоятельная работа	3 семестр - 93,5 часа;
в том числе на КП/КР	не предусмотрено учебным планом
Иная контактная работа	проводится в рамках часов аудиторных занятий
включая:	
Контрольная работа	
Промежуточная аттестация:	
Экзамен	3 семестр - 0,5 часа;

Москва 2024

ПРОГРАММУ СОСТАВИЛ:


Преподаватель

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Гданский Н.И.
	Идентификатор	Rb38b5a94-GdanskyNI-d2219177

Н.И. Гданский


СОГЛАСОВАНО:

Руководитель
образовательной
программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Варшавский П.Р.
	Идентификатор	R9a563c96-VarshavskyPR-efb4bbd

П.Р.
Варшавский

Заведующий выпускающей
кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Варшавский П.Р.
	Идентификатор	R9a563c96-VarshavskyPR-efb4bbd

П.Р.
Варшавский

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: изучение автоматизированных методов обработки больших данных в задачах искусственного интеллекта

Задачи дисциплины

- изучение методов построения регрессий;
- классификация регрессий;
- освоение языка статистических вычислений R;
- освоение методов построения регрессионных сплайнов;
- изучение методов обучения без учителя.

Формируемые у обучающегося **компетенции** и запланированные **результаты обучения** по дисциплине, соотнесенные с **индикаторами достижения компетенций**:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения
ПК-1 Способен выполнять работы на всем жизненном цикле информационных систем в выбранной среде разработки компьютерного ПО	ИД-2ПК-1 Демонстрирует знание современных программно-технических средств, информационных технологий и тенденции их развития	знать: - характеристики основных программно-технических сред для разработки проектов по информационным технологиям и методы работы в программно-технических средах для разработки проектов по информационным технологиям; - основные этапы жизненного цикла информационных систем в выбранной программно-технической среде и методы реализации разработки ПО в выбранной программно-технической среде. уметь: - осуществлять обоснованный выбор программно-технических сред для разработки проектов по информационным технологиям и практически применять программно-технические среды для разработки проектов по информационным технологиям; - разрабатывать ПО в выбранной программно-технической среде.
ПК-3 Способен применять методы и технологии искусственного интеллекта для разработки интеллектуальных систем	ИД-2ПК-3 Выбирает методы и средства обработки больших данных	знать: - методы и средства обработки больших данных. уметь: - применять для решения практических задач методы и средства обработки больших данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВО

Дисциплина относится к основной профессиональной образовательной программе Искусственный интеллект (далее – ОПОП), направления подготовки 01.04.02 Прикладная математика и информатика, уровень образования: высшее образование - магистратура.

Требования к входным знаниям и умениям:

- знать Основы программирования Основы теории вероятности и статистики Сетевые технологии

- уметь Программировать Строить регрессионные зависимости Рассчитывать параметры регрессий Практически использовать сетевые технологии

Результаты обучения, полученные при освоении дисциплины, необходимы при выполнении выпускной квалификационной работы.

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 144 часа.

№ п/п	Разделы/темы дисциплины/формы промежуточной аттестации	Всего часов на раздел	Семестр	Распределение трудоемкости раздела (в часах) по видам учебной работы										Содержание самостоятельной работы/ методические указания
				Контактная работа							СР			
				Лек	Лаб	Пр	Консультация		ИКР		ПА	Работа в семестре	Подготовка к аттестации /контроль	
КПР	ГК	ИККП	ТК											
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Основные проблемы обработки больших данных	27	3	4	-	8	-	-	-	-	-	15	-	<u>Подготовка к текущему контролю:</u> Повторение материала по разделу ". Основные проблемы обработки больших данных" <u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу ". Основные проблемы обработки больших данных"
1.1	Основные проблемы обработки больших данных	27		4	-	8	-	-	-	-	-	15	-	
2	Основные виды регрессий. Классификация регрессий	27		4	-	8	-	-	-	-	-	15	-	<u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу "Основные виды регрессий. Классификация регрессий" <u>Подготовка к текущему контролю:</u> Повторение материала по разделу "Основные виды регрессий. Классификация регрессий" <u>Изучение материалов литературных источников:</u> [2], 20-40
2.1	Основные виды регрессий. Классификация регрессий	27		4	-	8	-	-	-	-	-	15	-	
3	Деревья решений, регрессионные деревья, деревья классификации	27		4	-	8	-	-	-	-	-	15	-	<u>Подготовка к текущему контролю:</u> Повторение материала по разделу "Деревья решений, регрессионные деревья, деревья классификации" <u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу
3.1	Деревья решений, регрессионные деревья, деревья	27	4	-	8	-	-	-	-	-	15	-		

	классификации												"Деревья решений, регрессионные деревья, деревья классификации" <u>Изучение материалов литературных источников:</u> [3], 10-30
4	Обучение без учителя. Методы кластеризации	27	4	-	8	-	-	-	-	-	15	-	<u>Подготовка к текущему контролю:</u> Повторение материала по разделу "Обучение без учителя. Методы кластеризации"
4.1	Обучение без учителя. Методы кластеризации	27	4	-	8	-	-	-	-	-	15	-	<u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу "Обучение без учителя. Методы кластеризации" <u>Изучение материалов литературных источников:</u> [4], 30-50
	Экзамен	36.0	-	-	-	-	2	-	-	0.5	-	33.5	
	Всего за семестр	144.0	16	-	32	-	2	-	-	0.5	60	33.5	
	Итого за семестр	144.0	16	-	32		2		-	0.5		93.5	

Примечание: Лек – лекции; Лаб – лабораторные работы; Пр – практические занятия; КПр – аудиторные консультации по курсовым проектам/работам; ИККП – индивидуальные консультации по курсовым проектам/работам; ГК- групповые консультации по разделам дисциплины; СР – самостоятельная работа студента; ИКР – иная контактная работа; ТК – текущий контроль; ПА – промежуточная аттестация

3.2 Краткое содержание разделов

1. Основные проблемы обработки больших данных

1.1. Основные проблемы обработки больших данных

Статистический вывод. Задача регрессионного анализа. Разнородность. Накопление шума. Разреженные модели. Ложная корреляция. Случайная эндогенность.

2. Основные виды регрессий. Классификация регрессий

2.1. Основные виды регрессий. Классификация регрессий

Линейная регрессия. Логистическая регрессия. Гребневая регрессия (ридж-регрессия). Лассо-регрессия. Экологическая регрессия. Байесовская регрессия. Логическая регрессия. Квантильная регрессия. LAD-регрессия. Джекнайф-регрессия. Дискриминантный анализ. Язык статистических вычислений R, основные команды, работа с данными..

3. Деревья решений, регрессионные деревья, деревья классификации

3.1. Деревья решений, регрессионные деревья, деревья классификации

Описание структуры деревьев решений. Способы проверки условий. Построение деревьев на основе рекурсивного разбиения. Построение деревьев с использованием алгоритма условного вывода..

4. Обучение без учителя. Методы кластеризации

4.1. Обучение без учителя. Методы кластеризации

Процедура обучения. Управление набором данных. Отсечение ребер. Конгломерация и усиление. Цели кластеризации. Постановка задачи кластеризации. Типы кластерных структур. Методы решения. Графовые методы кластеризации. Иерархическая кластеризация (таксономия). Статистические методы кластеризации..

3.3. Темы практических занятий

1. Основные проблемы обработки больших данных;
2. Обучение без учителя. Методы кластеризации;
3. Деревья решений, регрессионные деревья, деревья классификации;
4. Основные виды регрессий..

3.4. Темы лабораторных работ

не предусмотрено

3.5 Консультации

Групповые консультации по разделам дисциплины (ГК)

1. Обсуждение материалов по кейсам раздела ". Основные проблемы обработки больших данных"
2. Обсуждение материалов по кейсам раздела "Основные виды регрессий. Классификация регрессий"
3. Обсуждение материалов по кейсам раздела "3. Деревья решений, регрессионные деревья, деревья классификации"
4. Обсуждение материалов по кейсам раздела "Обучение без учителя. Методы кластеризации"

3.6 Тематика курсовых проектов/курсовых работ

Курсовой проект/ работа не предусмотрены

3.7. Соответствие разделов дисциплины и формируемых в них компетенций

Запланированные результаты обучения по дисциплине (в соответствии с разделом 1)	Коды индикаторов	Номер раздела дисциплины (в соответствии с п.3.1)				Оценочное средство (тип и наименование)
		1	2	3	4	
Знать:						
основные этапы жизненного цикла информационных систем в выбранной программно-технической среде и методы реализации разработки ПО в выбранной программно-технической среде	ИД-2ПК-1	+				Контрольная работа/Основные проблемы обработки больших данных
характеристики основных программно-технических сред для разработки проектов по информационным технологиям и методы работы в программно-технических средах для разработки проектов по информационным технологиям	ИД-2ПК-1		+	+		Контрольная работа/Деревья решений, регрессионные деревья, деревья классификации Контрольная работа/Основные виды регрессий. Классификация регрессий
методы и средства обработки больших данных	ИД-2ПК-3				+	Контрольная работа/Обучение без учителя. Методы кластеризации
Уметь:						
разрабатывать ПО в выбранной программно-технической среде	ИД-2ПК-1		+	+		Контрольная работа/Деревья решений, регрессионные деревья, деревья классификации Контрольная работа/Основные виды регрессий. Классификация регрессий
осуществлять обоснованный выбор программно-технических сред для разработки проектов по информационным технологиям и практически применять программно-технические среды для разработки проектов по информационным технологиям	ИД-2ПК-1	+				Контрольная работа/Основные проблемы обработки больших данных
применять для решения практических задач методы и средства обработки больших данных	ИД-2ПК-3				+	Контрольная работа/Обучение без учителя. Методы

						кластеризации
--	--	--	--	--	--	---------------

4. КОМПЕТЕНТНОСТНО-ОРИЕНТИРОВАННЫЕ ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ)

4.1. Текущий контроль успеваемости

3 семестр

Форма реализации: Компьютерное задание

1. Деревья решений, регрессионные деревья, деревья классификации (Контрольная работа)

Форма реализации: Письменная работа

1. Обучение без учителя. Методы кластеризации (Контрольная работа)
2. Основные виды регрессий. Классификация регрессий (Контрольная работа)
3. Основные проблемы обработки больших данных (Контрольная работа)

Балльно-рейтинговая структура дисциплины является приложением А.

4.2 Промежуточная аттестация по дисциплине

Экзамен (Семестр №3)

Оценка определяется в соответствии с Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ» на основании семестровой и экзаменационной составляющих.

В диплом выставляется оценка за 3 семестр.

Примечание: Оценочные материалы по дисциплине приведены в фонде оценочных материалов ОПОП.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Печатные и электронные издания:

1. А.М. Годин- "Статистика", (11-е изд., перераб. и испр.), Издательство: "Издательско-торговая корпорация «Дашков и К°»", Москва, 2017 - (412 с.)
<http://biblioclub.ru/index.php?page=book&id=452543>;
2. Подольский, А. И. Молодежь и интернет-контент: анализ больших данных / А. И. Подольский, В. Л. Гойко // Вопросы психологии . – 2021 . – №3 . – С. 46-54 .;
3. Адлер Ю. П., Черных Е. А.- "Статистическое управление процессами. «Большие данные»", Издательство: "МИСИС", Москва, 2016 - (52 с.)
<https://e.lanbook.com/book/93652>;
4. Просто о больших данных : пер. с англ. / Д. Гурвиц, и др. – М. : Эксмо, 2015 . – 400 с. – (Библиотека Сбербанка ; Т.58) . - ISBN 978-5-699-85807-1 ..

5.2 Лицензионное и свободно распространяемое программное обеспечение:

1. СДО "Прометей";
2. Office / Российский пакет офисных программ;
3. Windows / Операционная система семейства Linux;
4. Видеоконференции (Майнд, Сберджаз, ВК и др);
5. Visual Studio;
6. Python.

5.3 Интернет-ресурсы, включая профессиональные базы данных и информационно-справочные системы:

1. ЭБС Лань - <https://e.lanbook.com/>
2. ЭБС "Университетская библиотека онлайн" - http://biblioclub.ru/index.php?page=main_ub_red
3. Электронная библиотека МЭИ (ЭБ МЭИ) - <http://elib.mpei.ru/login.php>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Тип помещения	Номер аудитории, наименование	Оснащение
Учебные аудитории для проведения лекционных занятий и текущего контроля	Ж-120, Машинный зал ИВЦ	сервер, кондиционер
	Г-401, Учебная аудитория	парта со скамьей, стол преподавателя, стул, доска меловая, компьютерная сеть с выходом в Интернет
Учебные аудитории для проведения практических занятий, КР и КП	М-708, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
	М-706, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
Учебные аудитории для проведения промежуточной аттестации	М-708, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
	М-706, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
Помещения для самостоятельной работы	НТБ-201, Компьютерный читальный зал	стол компьютерный, стул, стол письменный, вешалка для одежды, компьютерная сеть с выходом в Интернет, компьютер персональный, принтер, кондиционер
Помещения для консультирования	М-706а, Консультационный зал кафедры ПМИИ	парта со скамьей, стол, стул, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор
	М-704, Преподавательская кафедра ПМИИ	стол, стул, шкаф, тумба, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, холодильник, кондиционер
Помещения для хранения оборудования и учебного инвентаря	Е-403, Склад	стол для работы с документами, шкаф, шкаф для документов, книги, учебники, пособия, дипломные и курсовые работы студентов

БАЛЛЬНО-РЕЙТИНГОВАЯ СТРУКТУРА ДИСЦИПЛИНЫ**Технологии обработки больших данных**

(название дисциплины)

3 семестр**Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:**

- КМ-1 Основные проблемы обработки больших данных (Контрольная работа)
 КМ-2 Основные виды регрессий. Классификация регрессий (Контрольная работа)
 КМ-3 Деревья решений, регрессионные деревья, деревья классификации (Контрольная работа)
 КМ-4 Обучение без учителя. Методы кластеризации (Контрольная работа)

Вид промежуточной аттестации – Экзамен.

Номер раздела	Раздел дисциплины	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
		Неделя КМ:	4	8	12	16
1	Основные проблемы обработки больших данных					
1.1	Основные проблемы обработки больших данных		+			
2	Основные виды регрессий. Классификация регрессий					
2.1	Основные виды регрессий. Классификация регрессий			+	+	
3	Деревья решений, регрессионные деревья, деревья классификации					
3.1	Деревья решений, регрессионные деревья, деревья классификации			+	+	
4	Обучение без учителя. Методы кластеризации					
4.1	Обучение без учителя. Методы кластеризации					+
Вес КМ, %:			25	25	25	25