

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по безопасности и общим вопросам

Дата подписания: 27.04.2023 16:31:20

Уникальный программный ключ:

d7a26b9e8ca85e98ac3de2ab454b4659d961f749

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский технологический университет «МИСиС»

Рабочая программа дисциплины (модуля)

Машинное обучение

Закреплена за подразделением

Кафедра теоретической физики и квантовых технологий

Направление подготовки

03.03.02 ФИЗИКА

Профиль

Квалификация

Бакалавр

Форма обучения

очная

Общая трудоемкость

4 ЗЕТ

Часов по учебному плану

144

Формы контроля в семестрах:

в том числе:

зачет с оценкой 6

аудиторные занятия

51

самостоятельная работа

93

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	6 (3.2)		Итого	
	18			
Неделя	УП	РП	УП	РП
Лекции	17	17	17	17
Практические	34	34	34	34
Итого ауд.	51	51	51	51
Контактная работа	51	51	51	51
Сам. работа	93	93	93	93
Итого	144	144	144	144

Программу составил(и):

к.ф.-.м.н., доц., Смирнова Екатерина Александровна; к.ф.-.м.н., доц., Галимзянов Тимур Равильевич

Рабочая программа

Машинное обучение

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - бакалавриат Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСиС» по направлению подготовки 03.03.02 ФИЗИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

03.03.02 ФИЗИКА, 03.03.02-БФ3-22.plx , утвержденного Ученым советом ФГАОУ ВО НИТУ "МИСиС" в составе соответствующей ОПОП ВО 22.09.2022, протокол № 8-22

Утверждена в составе ОПОП ВО:

03.03.02 ФИЗИКА, , утвержденной Ученым советом ФГАОУ ВО НИТУ "МИСиС" 22.09.2022, протокол № 8-22

Рабочая программа одобрена на заседании

Кафедра теоретической физики и квантовых технологий

Протокол от г., №

Руководитель подразделения Мухин Сергей Иванович, д.ф.-м.н., профессор

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Цель дисциплины - ознакомление студентов с основами и принципами машинного обучения — а именно, с классами моделей (линейные, логические, нейросетевые), метриками качествами и подходами к подготовке данных.
-----	--

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В.ДВ.04
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Линейная алгебра	
2.1.2	Методы контроля и анализа веществ	
2.1.3	Техника физико-химического эксперимента	
2.1.4	Кристаллография	
2.1.5	Физика	
2.1.6	Физическая химия	
2.1.7	Химия	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.2	Преддипломная практика для выполнения выпускной квалификационной работы	
2.2.3	Теоретическая нанофотоника	
2.2.4	Методы физико-химических исследований	
2.2.5	Оформление результатов научной деятельности	
2.2.6	Специальный физический практикум	
2.2.7	Методы вычислительной физики	
2.2.8	Нормы и правила оформления ВКР	
2.2.9	Преддипломная практика для выполнения выпускной квалификационной работы	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ОПК-2: Способен проводить научные исследования физических объектов, систем и процессов, обрабатывать и представлять экспериментальные данные, демонстрировать навыки работы в лаборатории / мастерской, способность разрабатывать и проводить экспериментальные исследования, интерпретировать данные и делать выводы в соответствующей области исследования	
Знать:	
ОПК-2-32	Способы формализации признаковового описания объектов
ОПК-2-31	Основные типы алгоритмов машинного обучения и особенности их применения
Уметь:	
ОПК-2-У1	Решать задачи кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных
Владеть:	
ОПК-2-В1	навыками использования современных свободно распространяемых программных средств для машинного обучения, визуализации, интерпретации и анализа данных

4. СТРУКТУРА И СОДЕРЖАНИЕ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Формируемые индикаторы компетенций	Литература и эл. ресурсы	Примечание	КМ	Выполняемые работы
	Раздел 1. Введение. Типы задач. Метрические классификаторы. Алгоритмы кластеризации							

1.1	Необходимые понятия из линейной алгебры, математического анализа, теории вероятностей и численной оптимизации. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.2	Типы задач машинного обучения. Обзор задач, решаемых алгоритмами машинного обучения. Классификация алгоритмов машинного обучения. Обучающая, тестовая и валидационная выборка. Недообучение и переобучение. Параметры и гиперпараметры. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.3	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия. Регуляризация. Логистическая регрессия. Целевая функция логистической регрессии. Регуляризация логистической регрессии. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.4	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия. Регуляризация. Логистическая регрессия. Целевая функция логистической регрессии. Регуляризация логистической регрессии. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.5	Общий вид метрического классификатора. Алгоритм К ближайших соседей. Алгоритмы отбора эталонов. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.6	Алгоритм К ближайших соседей. Алгоритмы отбора эталонов. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.7	Алгоритмы кластеризации с фиксированным количеством кластеров. Алгоритмы кластеризации по плотности. Иерархическая кластеризация. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
1.8	Освоение теоретического материала /Ср/	6	25	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
	Раздел 2. Деревья решений. Нейронные сети							

2.1	Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.2	Алгоритм построения деревьев решений. Леса решающих деревьев. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.3	Структура нейрона. Структура нейронной сети. Градиентный спуск. Стохастический градиентный спуск. Обучения нейронной сети с помощью алгоритма обратного распространения ошибки. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.4	L1 и L2-регуляризация весов. Dropout. Аугментация выборки. Батч-нормализация. Instance-нормализация. Cyclic learning rate. Fine-tuning. /Лек/	6	1	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.5	Глубокое обучение, Сверточные нейронные сети, пулинг, внимание. Residual connections. Рекуррентные нейронные сети, трансформеры, GRU, LSTM. Двусторонние рекуррентные нейронные сети. Алгоритм ВРГТ. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.6	Автоэнкодеры. Вариационные автоэнкодеры. Генеративно-состязательные сети. Adversarial autoencoders. Идентификация личности. Triplet loss. Машинный перевод. Seq2seq. Attention, трансформеры. /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
2.7	Практическое занятие. Функции активации. /Пр/	6	4	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1 Э1			Р1
2.8	Практическое занятие. Инициализация и оптимизация нейросетей /Пр/	6	4	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1 Э1			Р3
2.9	Практическое занятие. Сверточные нейросети. /Пр/	6	4	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1 Э1			
2.10	Освоение теоретического и практического материала. /Ср/	6	25	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			

	Раздел 3. Ансамблевые методы. Стохастический поиск							
3.1	Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг. Алгоритм Random Forest. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
3.2	Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм. /Лек/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
3.3	Генетический алгоритм /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1			
3.4	Практическое занятие по глубокому обучению. /Пр/	6	4	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1 Э1			
3.5	Коллоквиум /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1		КМ1	
3.6	Выполнение проекта /Ср/	6	43	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1 Э1			
3.7	Защита проекта /Пр/	6	2	ОПК-2-31 ОПК-2-32 ОПК-2-У1 ОПК-2-В1	Л1.1Л2.1		КМ2	

5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

5.1. Контрольные мероприятия (контрольная работа, тест, коллоквиум, экзамен и т.п), вопросы для самостоятельной подготовки

Код КМ	Контрольное мероприятие	Проверяемые индикаторы компетенций	Вопросы для подготовки
--------	-------------------------	------------------------------------	------------------------

КМ1	Коллоквиум	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	<ol style="list-style-type: none"> 1. Теорема Байеса. Общая формулировка, примеры применения. 2. Метод максимального правдоподобия. Общая формулировка, примеры применения. 3. Работа с данными в машинном обучении - обучающая, тестовая и валидационная выборки. Недообучение и переобучение. Параметры и гиперпараметры. Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. 4. Линейная и полиномиальные регрессии. Описание метода, области применения. Регуляризация. 5. Логистическая регрессия. Описание метода, области применения. Регуляризация. 6. Описание основных алгоритмов метрических классификаторов. 7. Описание основных алгоритмов метрических кластеризации. 8. Решающие деревья. Алгоритм построения. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев. 9. Структура нейрона. Структура нейронной сети. Градиентный спуск. Стохастический градиентный спуск. Обучение нейронной сети с помощью алгоритма обратного распространения ошибки. 10. Методы регуляризации нейронных сетей. 11. Глубокое обучение. Сверточные нейронные сети. 12. Глубокое обучение. Механизм внимания, трансформеры. 13. Рекуррентные нейронные сети. 14. Автоэнкодеры. Вариационные автоэнкодеры. 15. Идентификация личности. Triplet loss. 16. Ансамблевые методы машинного обучения. 17. Алгоритмы стохастического поиска.
КМ2	Защита проекта	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	<ol style="list-style-type: none"> 1. Обоснование выбора метода 2. Принципы построения модели. 3. Точность полученных результатов

5.2. Перечень работ, выполняемых по дисциплине (Курсовая работа, Курсовой проект, РГР, Реферат, ЛР, ПР и т.п.)

Код работы	Название работы	Проверяемые индикаторы компетенций	Содержание работы
P1	Практическое занятие 1	ОПК-2-31;ОПК-2-У1;ОПК-2-32;ОПК-2-В1	Вариант в приложении. Функции активации.
P2	Практическое занятие 2	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	Вариант в приложении. Сверточные нейросети.
P3	Практическое занятие 3	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	Вариант в приложении. Оптимизация и инициализация нейронных сетей.
P4	Практическое занятие 4	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	Вариант в приложении. Использование машинного обучения для физических задач

P5	Проект	ОПК-2-31;ОПК-2-32;ОПК-2-У1;ОПК-2-В1	<p>Варианты проектов</p> <p>1. Предсказание извержения вулкана по данным сейсмических сенсоров.</p> <p>Цель проекта – освоение принципов работы с табличными данными большого объёма в практических физических (геофизических) задачах. В данном проекте предлагается построить прогностическую модель извержения вулканов на основе анализа большого набора геофизических данных, собранных датчиками, развернутыми на действующих вулканах.</p> <p>Входные данные представляют собой сигналы сейсмических сенсоров. Каждый файл содержит 10-минутную запись 10 сенсоров вокруг каждого вулкана. Целевой переменной является время до следующего извержения</p> <p>Ссылка на данные: https://www.kaggle.com/competitions/predict-volcanic-eruptions-ingv-oe/data.</p> <p>2. Построение модели обнаружения дефектов в образцах стали по микрофотографии. Цель проекта – освоение методов обработки изображений для анализа качества продукции промышленных предприятий. В данном проекте предполагается построить модель, позволяющую по микрофотографии локализовать и классифицировать поверхностные дефекты стального листа.</p> <p>Входные данные представляют собой микрофотографии. Целевой переменной является метка дефекта.</p> <p>Ссылка на данные: https://www.kaggle.com/competitions/severstal-steel-defect-detection/data</p> <p>3. Анализ треков элементарных частиц в детекторах ускорителей элементарных частиц. Цель проекта – освоение принципов работы с табличными данными большого объёма в физических задачах. В данном проекте предполагается создать модель, которая реконструирует треки частиц из трехмерных точек, оставленных в кремниевых детекторах.</p> <p>Входные данные представляют собой данные о координатах точки взаимодействия частицы с детектором, идентификатор детектора и его положение. Целевой переменной является идентификатор трека частицы.</p> <p>Ссылка на данные: https://www.kaggle.com/competitions/trackml-particle-identification/data</p> <p>4. Анализ изображений для распознавания архитектурных и географических объектов. Цель проекта – освоение методов обработки изображений для практических задач классификации. В ходе проекта необходимо будет построить модель, которая по данным изображениям распознаёт архитектурный или географический объект. Особенностью задания является большое число классов в обучающем наборе данных (около 15 000) и малое количество обучающих примеров на класс.</p> <p>Входные данные представляют собой набор фотографий, на которых присутствуют целевые объекты. Целевой переменной является идентификатор объекта.</p> <p>Ссылка на данные: https://www.kaggle.com/competitions/landmark-recognition-challenge/overview</p> <p>5. Анализ настроений в коротких текстах. Цель проекта –</p>
----	--------	-------------------------------------	--

			<p>освоение анализа текстовых данных в задаче классификации. В данном проекте предполагается построить модель, которая позволяет по короткому сообщению определить его эмоциональную окраску.</p> <p>Входные данные представляют собой короткие тексты. Целевой переменной является метка эмоциональной окраски.</p> <p>Ссылка на данные: https://www.kaggle.com/competitions/tweet-sentiment-extraction/data</p>
--	--	--	---

5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)

5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Северенс Ч.	Введение в программирование на Python	Электронная библиотека	Москва: Национальный Открытый Университет «ИНТУИТ», 2016

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Google Colab	https://colab.research.google.com/
----	--------------	---

6.3 Перечень программного обеспечения

П.1	Лицензии ПО Windows Server CAL ALNG LicSAPk MVL DvcCAL, ПО WinEDUA3 ALNG SubsVL MVL PerUsr и PerUsr
П.2	Microsoft Office
П.3	MS Teams
П.4	Python
П.5	ОС Linux (Ubuntu) / Windows
П.6	Putty

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Полнотекстовые российские научные журналы и статьи:
И.2	— Научная электронная библиотека eLIBRARY https://elibrary.ru/
И.3	Иностранные базы данных (доступ с IP адресов МИСиС):
И.4	— аналитическая база (индексы цитирования) Web of Science https://apps.webofknowledge.com
И.5	— аналитическая база (индексы цитирования) Scopus https://www.scopus.com/
И.6	— наукометрическая система InCites https://apps.webofknowledge.com
И.7	— научные журналы издательства Elsevier https://www.sciencedirect.com/

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

Дисциплина относится к точным наукам и требует значительного объема самостоятельной работы. Отдельные учебные вопросы выносятся на самостоятельную проработку и контролируются посредством текущей аттестации. При этом организуются групповые и индивидуальные консультации. Качественное освоение дисциплины возможно только при систематической самостоятельной работе.