

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ
КАФЕДРА КИБЕРНЕТИКИ

ОДОБРЕНО УМС ИИКС

Протокол № 8/1/2024

от 28.08.2024 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ

Направление подготовки
(специальность)

[1] 01.04.02 Прикладная математика и информатика

Семестр	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	В форме практической подготовки/ В	СРС, час.	КСР, час.	Форма(ы) контроля, экс./зач./КР/КП
1	4	144	32	16	16		44	0	Э
Итого	4	144	32	16	16	0	44	0	

АННОТАЦИЯ

Машинное обучение – раздел искусственного интеллекта, изучающий методы построения моделей и алгоритмов, способных обучаться. Методы машинного обучения используются при решении широкого круга прикладных задач, для которых разработка явного алгоритма решения затруднительна или невозможна. И этот круг задач постоянно расширяется. Повсеместная информатизация приводит к накоплению огромных объемов данных в науке, производстве, бизнесе, транспорте, здравоохранении.

В настоящем курсе рассматриваются основные принципы машинного обучения с учителем, изучаются постановки задач регрессии и классификации и используемые для их решения модели и методы обучения, а также оценки качества обученных моделей, рассматриваются особенности организации процесса обучения с учителем и применения методов машинного обучения для решения практических задач. Изложение ведется строгим математическим языком, сопровождается множеством формул и математических выкладок.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью учебного курса является ознакомление будущих специалистов с основными принципами и современными математическими и алгоритмическими подходами к построению моделей машинного обучения.

Учебный курс ставит также своей целью приобретение умений ставить практические задачи в терминах машинного обучения, освоение студентами практических приемов построения моделей машинного обучения и оценки их качества. Выполняемое студентами домашнее задание позволит получить практический опыт по созданию и экспериментальному исследованию свойств обучающихся моделей и методов обучения.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина «Введение в машинное обучение» является обязательной для магистранта. Дисциплина требует от слушателя общематематической подготовки по дискретной математике, математическому анализу, линейной алгебре, методам оптимизации, а также углубленного знания теории вероятностей и математической статистики. Полученные знания могут быть применены студентами в процессе обучения при выполнении текущей научно-исследовательской и выпускной работ.

Дисциплина должна изучаться после прохождения курса «Математическая статистика». В свою очередь, дисциплина является предшествующей и необходимой для изучения следующих курсов профессионального цикла подготовки:

- Анализ данных и машинное обучение: часть 1;
- Машинное зрение.

Дисциплина формирует систему базовых понятий, необходимых для специалиста в области программной инженерии и больших данных, прикладной математики и информатики, способствует выработке профессиональных решений практических задач.

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные и(или) общепрофессиональные компетенции:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
УК-1 [1] – Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	З-УК-1 [1] – Знать: методы системного и критического анализа; методики разработки стратегии действий для выявления и решения проблемной ситуации У-УК-1 [1] – Уметь: применять методы системного подхода и критического анализа проблемных ситуаций; разрабатывать стратегию действий, принимать конкретные решения для ее реализации В-УК-1 [1] – Владеть: методологией системного и критического анализа проблемных ситуаций; методиками постановки цели, определения способов ее достижения, разработки стратегий действий
УК-6 [1] – Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	З-УК-6 [1] – Знать: методики самооценки, самоконтроля и саморазвития с использованием подходов здоровьесбережения У-УК-6 [1] – Уметь: решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности; применять методики самооценки и самоконтроля; применять методики, позволяющие улучшить и сохранить здоровье в процессе жизнедеятельности В-УК-6 [1] – Владеть: технологиями и навыками управления своей познавательной деятельностью и ее совершенствования на основе самооценки, самоконтроля и принципов самообразования в течение всей жизни, в том числе с использованием здоровьесберегающих подходов и методик

Профессиональные компетенции в соответствии с задачами и объектами (областями знаний) профессиональной деятельности:

Задача профессиональной деятельности (ЗПД)	Объект или область знания	Код и наименование профессиональной компетенции; Основание (профессиональный стандарт-ПС, анализ опыта)	Код и наименование индикатора достижения профессиональной компетенции
научно-исследовательский			
анализ и математическое моделирование динамических систем	сложные технические системы	ПК-8.1 [1] - способен разрабатывать математическое обеспечение киберфизических систем	З-ПК-8.1[1] - знать методы прикладной математики, используемые при построении математических

		<p><i>Основание:</i> Профессиональный стандарт: 32.002</p>	<p>моделей киберфизических систем; У-ПК-8.1[1] - уметь обоснованно выбирать и применять методы прикладной математики при разработке математического обеспечения киберфизических систем; В-ПК-8.1[1] - владеть программными средствами, используемыми при разработке математического обеспечения киберфизических систем</p>
<p>анализ и математическое моделирование физических процессов</p>	<p>системы ядерно-энергетического комплекса</p>	<p>ПК-1 [1] - способен проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива</p> <p><i>Основание:</i> Профессиональный стандарт: 24.078</p>	<p>З-ПК-1[1] - Знать основные методы и принципы научных исследований, математического моделирования, основные проблемы профессиональной области, требующие использования современных научных методов исследования. ; У-ПК-1[1] - Уметь ставить и решать прикладные исследовательские задачи; оценивать результаты исследований; формулировать результаты проведенного исследования в виде конкретных рекомендаций, проводить научные исследования и получать новые научные и</p>

			<p>прикладные результаты самостоятельно и в составе научного коллектива. ; В-ПК-1[1] - Владеть навыками выбора и использования математических средств научных исследований, методами анализа и синтеза научной информации.</p>
<p>анализ и математическое моделирование физических процессов</p>	<p>системы ядерно-энергетического комплекса</p>	<p>ПК-3 [1] - способен развивать инновационный потенциал новых научных и научно-технологических разработок</p> <p><i>Основание:</i> Профессиональный стандарт: 24.078</p>	<p>3-ПК-3[1] - Знать основы планирования и организации научных исследований в профессиональной области; методику постановки задач по решению теоретических и прикладных исследовательских проблем; методы и средства научных исследований в профессиональной области, правила и принципы научной этики, методы математического моделирования. ; У-ПК-3[1] - Уметь оценивать и развивать инновационный потенциал новых научных и научно-технологических разработок, осуществлять постановку задач по решению теоретических и прикладных исследовательских проблем; составить план научных исследований; выдвинуть гипотезы</p>

			<p>по направлению исследований и соотнести их с полученными результатами; организовать свою научно-исследовательскую работу; определять методы и средства научных исследований для решения конкретных задач в своей предметной области; оценивать результаты исследований, использовать методы математического моделирования;</p> <p>В-ПК-3[1] - Владеть навыками постановки задач по решению теоретических и прикладных исследовательских проблем; навыками выбора и использования методов и средств научных исследований задач в своей предметной области; навыками методами работы с литературными источниками; методами анализа результатов научных исследований; методами обобщения результатов научных исследований для развития инновационного потенциала новых научных и научно-технологических разработок</p>
анализ и математическое моделирование	системы ядерно-энергетического комплекса	ПК-4 [1] - способен проводить экспертизы инновационных	З-ПК-4[1] - Знать основные методы и принципы экспертизы

физических процессов		<p>проектов в сфере своей профессиональной деятельности</p> <p><i>Основание:</i> Профессиональный стандарт: 24.078</p>	<p>инновационных проектов в сфере своей профессиональной деятельности. ; У-ПК-4[1] - Уметь проводить экспертизы инновационных проектов, оценивать перспективы развития проектов в сфере своей профессиональной деятельности.; В-ПК-4[1] - Владеть навыками проведения экспертизы инновационных проектов в сфере своей профессиональной деятельности.</p>
производственно-технологический			
разработка и сопровождение программного обеспечения;	информационные и программные системы	<p>ПК-6 [1] - способен к проектированию и разработке наукоемкого программного обеспечения на основе технического задания</p> <p><i>Основание:</i> Профессиональный стандарт: 06.003</p>	<p>3-ПК-6[1] - Знать основные цели и задачи проектирования и разработки наукоемкого программного обеспечения на основе технического задания. ; У-ПК-6[1] - Уметь разрабатывать наукоемкое программное обеспечение на основе технического задания.; В-ПК-6[1] - Владеть навыками разработки и проектирования наукоемкого программного обеспечения на основе технического задания.</p>

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Разделы учебной дисциплины, их объем, сроки изучения и формы контроля:

№ п.п	Наименование раздела учебной дисциплины	Недели	Лекции/ Практи. (семинары)/ Лабораторные работы, час.	Обязат. текущий контроль (форма*, неделя)	Максимальный балл за раздел**	Аттестация раздела (форма*, неделя)	Индикаторы освоения компетенции
	<i>1 Семестр</i>						
1	Обучение с учителем: основные принципы	1-8	16/8/8		25	КИ-8	3-ПК-8.1, У-ПК-8.1, В-ПК-8.1, 3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-3, У-ПК-3, В-ПК-3, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-6, У-ПК-6, В-ПК-6, 3-УК-1, У-УК-1, В-УК-1, 3-УК-6, У-УК-6, В-УК-6
2	Обучение с учителем: классификация и регрессия	9-16	16/8/8		25	КИ-16	3-ПК-8.1, У-ПК-8.1, В-ПК-8.1, 3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-3, У-ПК-3, В-ПК-3, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-6, У-ПК-6, В-ПК-6, 3-УК-1, У-УК-1, В-УК-1, 3-УК-6, У-УК-6, В-УК-6
	<i>Итого за 1 Семестр</i>		32/16/16		50		

	Контрольные мероприятия за 1 Семестр				50	Э	3-ПК-8.1, У-ПК-8.1, В-ПК-8.1, 3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-3, У-ПК-3, В-ПК-3, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-6, У-ПК-6, В-ПК-6, 3-УК-1, У-УК-1, В-УК-1, 3-УК-6, У-УК-6, В-УК-6
--	---------------------------------------------	--	--	--	----	---	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

* – сокращенное наименование формы контроля

** – сумма максимальных баллов должна быть равна 100 за семестр, включая зачет и (или) экзамен

Сокращение наименований форм текущего контроля и аттестации разделов:

Обозначение	Полное наименование
КИ	Контроль по итогам
Э	Экзамен

КАЛЕНДАРНЫЙ ПЛАН

Недели	Темы занятий / Содержание	Лек., час.	Пр./сем., час.	Лаб., час.
	<i>1 Семестр</i>	32	16	16
1-8	Обучение с учителем: основные принципы	16	8	8
1 - 2	Основные парадигмы машинного обучения Машинное обучение и науки о данных. Data-driven и model-based подходы. История машинного обучения. Общая схема решения задач машинного обучения. Стандарт CRISP-DM. Типы задач машинного обучения. Примеры прикладных задач. Основные понятия и определения. Обзор основных парадигм машинного обучения.	Всего аудиторных часов		
		4	2	2
		Онлайн		
		0	0	0
3 - 5	Обучение с учителем: основные принципы Методы обучения с учителем. Обучающийся алгоритм. Индуктивный порог. Проблема переобучения и понятие обобщающей способности. Функция потерь. Теоретический и эмпирический риск. Принцип	Всего аудиторных часов		
		6	3	3
		Онлайн		
		0	0	0

	минимизации эмпирического риска. Смещение и дисперсия ошибки обучаемой модели. Разложение «bias-variance». Связь сложности модели, индуктивного порога и способности к обобщению. Оценивание точности обученных моделей. Методы кросс-валидации: hold-out, k-fold, LOOCV. Стратификация выборки при кросс-валидации, внутренняя кросс-валидация.			
6 - 8	Обучение с учителем: регрессия Понятие статистической модели. Теоретические и статистические модели. Примеры. Регрессионные модели. Постановка задачи обучения регрессионной модели. Риск модели с квадратичной функцией потерь. Функция регрессии. Оптимальность регрессионных моделей. Задачи, виды и этапы регрессионного анализа. Метод наименьших квадратов. Матрица плана и система нормальных уравнений. МНК-оценки параметров простейшей линейной регрессии. Ошибки и остатки модели. Предположения регрессионного анализа. Свойства МНК-оценок: линейность, состоятельность, несмещенность, эффективность, нормальность. Теорема Гаусса-Маркова. Метод наименьших квадратов. Проекционная матрица. МНК-оценки параметров множественной линейной регрессии. Статистические характеристики МНК-оценок. Проблема гетероскедастичности. Статистические тесты на гетероскедастичность. Визуальный анализ гетероскедастичности остатков регрессионной модели. Подходы к устранению гетероскедастичности. Взвешенный метод наименьших квадратов. Робастный регрессионный анализ. Метод наименьших модулей, LAR-регрессия. Постановка задачи и обучение. Использование взвешенного метода наименьших квадратов для обучения робастных моделей. Метод итеративного взвешивания (Iteratively Reweighted Least-Squares, IRLS).	Всего аудиторных часов		
		6	3	3
		Онлайн		
		0	0	0
9-16	Обучение с учителем: классификация и регрессия	16	8	8
9 - 11	Диагностика регрессионных моделей Цели и виды диагностики регрессионных моделей. Проверка линейности. Графики остатков и частичной регрессии. Смещение оценок вследствие неучтенных переменных (omitted variable bias). Проверка гетероскедастичности, независимости и нормальности регрессионных остатков. Визуальная и количественная диагностика. Квантильные диаграммы. Понятие регрессионного выброса. Показатели аномальности наблюдения. Диаграмма box-and-whisker. Оценка влияния наблюдений на регрессионную модель. Связь аномальности с леввериджами и стандартизованными регрессионными остатками. Расстояние Кука. Понятие мультиколлинеарности. Негативные эффекты мультиколлинеарности регрессоров. Показатели мультиколлинеарности. Variance inflation factor (VIF). Методы устранения мультиколлинеарности.	Всего аудиторных часов		
		6	3	3
		Онлайн		
		0	0	0

	Регуляризованные регрессионные модели. Гребневая регрессия. LASSO-регрессия. Подходы к выбору регрессоров. Пошаговая регрессия.			
12 - 14	Бинарная классификация Постановка задачи классификации. Бинарная классификация. Функция потерь и эмпирический риск классификатора. Виды функций потерь. Статистический взгляд на задачу обучения. Принцип минимизации эмпирического риска и метод максимального правдоподобия. Регуляризованный эмпирический риск. Оценка точности бинарного классификатора. Матрица ошибок. Специфичность и чувствительность, точность и полнота. ROC-анализ и PR-анализ. Показатели ROC AUC и PR AUC.	Всего аудиторных часов		
		6	3	3
		Онлайн		
		0	0	0
15 - 16	Многоклассовая классификация Постановка задачи многоклассовой классификации. Подходы к многоклассовой классификации. Метод ЕСОС. Схемы кодирования. Методы декодирования. Оценка точности многоклассового классификатора. Матрица ошибок. Микро- и макро- показатели точности. Специфичность и чувствительность, точность и полнота. Микро- и макро- ROC-анализы и PR-анализы.	Всего аудиторных часов		
		4	2	2
		Онлайн		
		0	0	0

Сокращенные наименования онлайн опций:

Обозначение	Полное наименование
ЭК	Электронный курс
ПМ	Полнотекстовый материал
ПЛ	Полнотекстовые лекции
ВМ	Видео-материалы
АМ	Аудио-материалы
Прз	Презентации
Т	Тесты
ЭСМ	Электронные справочные материалы
ИС	Интерактивный сайт

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

1. Лекционные занятия:

а. комплект электронных презентаций/слайдов,

б. аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук)

2. Практические и лабораторные занятия:

а. компьютерный класс,

б. презентационная техника (проектор, экран, компьютер/ноутбук),

с. выход в интернет.

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущего, рубежного и промежуточного контроля по дисциплине.

Связь между формируемыми компетенциями и формами контроля их освоения представлена в следующей таблице:

Компетенция	Индикаторы освоения	Аттестационное мероприятие (ЖП 1)
ПК-1	З-ПК-1	Э, КИ-8, КИ-16
	У-ПК-1	Э, КИ-8, КИ-16
	В-ПК-1	Э, КИ-8, КИ-16
ПК-3	З-ПК-3	Э, КИ-8, КИ-16
	У-ПК-3	Э, КИ-8, КИ-16
	В-ПК-3	Э, КИ-8, КИ-16
ПК-4	З-ПК-4	Э, КИ-8, КИ-16
	У-ПК-4	Э, КИ-8, КИ-16
	В-ПК-4	Э, КИ-8, КИ-16
ПК-6	З-ПК-6	Э, КИ-8, КИ-16
	У-ПК-6	Э, КИ-8, КИ-16
	В-ПК-6	Э, КИ-8, КИ-16
ПК-8.1	З-ПК-8.1	Э, КИ-8, КИ-16
	У-ПК-8.1	Э, КИ-8, КИ-16
	В-ПК-8.1	Э, КИ-8, КИ-16
УК-1	З-УК-1	Э, КИ-8, КИ-16
	У-УК-1	Э, КИ-8, КИ-16
	В-УК-1	Э, КИ-8, КИ-16
УК-6	З-УК-6	Э, КИ-8, КИ-16
	У-УК-6	Э, КИ-8, КИ-16
	В-УК-6	Э, КИ-8, КИ-16

Шкалы оценки образовательных достижений

Шкала каждого контрольного мероприятия лежит в пределах от 0 до установленного максимального балла включительно. Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет собой сумму баллов, заработанных студентом при выполнении заданий в рамках текущего и промежуточного контроля.

Итоговая оценка выставляется в соответствии со следующей шкалой:

Сумма баллов	Оценка по 4-х балльной шкале	Отметка о зачете	Оценка ECTS
90-100	5 – «отлично»	«Зачтено»	A
85-89			B
75-84			C
70-74			D
65-69			3 – «удовлетворительно»
60-64	F		
Ниже 60	2 – «неудовлетворительно»	«Не зачтено»	

Оценка «отлично» соответствует глубокому и прочному освоению материала программы обучающимся, который последовательно, четко и логически стройно излагает свои ответы, умеет тесно увязывать теорию с практикой, использует в ответах материалы монографической литературы.

Оценка «хорошо» соответствует твердым знаниям материала обучающимся, который грамотно и, по существу, излагает свои ответы, не допуская существенных неточностей.

Оценка «удовлетворительно» соответствует базовому уровню освоения материала обучающимся, при котором освоен основной материал, но не усвоены его детали, в ответах присутствуют неточности, недостаточно правильные формулировки, нарушения логической последовательности.

Отметка «зачтено» соответствует, как минимум, базовому уровню освоения материала программы, при котором обучающийся владеет необходимыми знаниями, умениями и навыками, умеет применять теоретические положения для решения типовых практических задач.

Оценку «неудовлетворительно» / отметку «не зачтено» получает обучающийся, который не знает значительной части материала программы, допускает в ответах существенные ошибки, не выполнил все обязательные задания, предусмотренные программой. Как правило, такие обучающиеся не могут продолжить обучение без дополнительных занятий.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ Г 77 Data Science. Наука о данных с нуля: Пер. с англ. — 2-е изд., перераб. и доп. : , Грас Д., Санкт-Петербург: БХВ-Петербург, 2021
2. ЭИ П 37 Python для сложных задач: наука о данных и машинное обучение : , Плас Дж. Вандер, Санкт-Петербург: Питер, 2018
3. ЭИ М 63 Введение в анализ данных : учебник и практикум, Миркин Б. Г., Москва: Юрайт, 2022

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. ЭИ Ф 70 Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : , Флах П., Москва: ДМК Пресс, 2015

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

Специальное программное обеспечение не требуется

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

<https://online.mephi.ru/>

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

9. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

В качестве оценочного средства используется 100 бальная семестровая система, учитывающая посещаемость занятий, активность (выполнение домашних занятий), выполнение тематических домашних заданий по каждому разделу, контрольно-тестовая работа по каждому разделу. Каждый раздел проходит аттестацию.

Итоговый балл за раздел (КИ) формируется следующим образом:
посещаемость семинарских занятий (еженедельно) не менее 80% +2 балла
не менее 50% +1 балл
менее 50% 0 баллов

ДЗ – выполнения тематического ДЗ (по каждому разделу)

Выполнено не менее 80% +3 баллов

Выполнено не менее 50% +2 балла

Выполнено менее 50% 0 баллов

Самостоятельная работа студента включает: Повторение теоретического материала –
Выполнение ДЗ

10. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

В качестве оценочного средства используется 100 бальная семестровая система, учитывающая посещаемость занятий, активность (выполнение домашних занятий), выполнение тематических домашних заданий по каждому разделу, контрольно-тестовая работа по каждому разделу. Каждый раздел проходит аттестацию.

Итоговый балл за раздел (КИ) формируется следующим образом:
посещаемость семинарских занятий (еженедельно) не менее 80% +2 балла
не менее 50% +1 балл
менее 50% 0 баллов

ДЗ – выполнения тематического ДЗ (по каждому разделу)

Выполнено не менее 80% +3 баллов

Выполнено не менее 50% +2 балла

Выполнено менее 50% 0 баллов

Самостоятельная работа студента включает: Повторение теоретического материала –
Выполнение ДЗ

Автор(ы):

Трофимов Александр Геннадьевич, к.т.н.