

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
АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ ЧАСТЬ 1

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
2	3	108	15	15	15		27	0	Э
Итого	3	108	15	15	15	0	27	0	

АННОТАЦИЯ

В курсе «Анализ данных и машинное обучение: часть 1» продолжают рассматриваться основные модели и методы машинного обучения. Дисциплина призвана обеспечить освоение студентами подходов к машинному обучению с учителем и без учителя, методов статистической интерпретации процесса обучения и анализа обученных моделей. Содержание дисциплины включает изучение байесовского подхода к обучению, генеративные и дискриминативные модели классификации, методы построения деревьев решений и ансамблей моделей машинного обучения, основы непараметрического анализа данных и обучения без учителя. Теоретические основы машинного обучения закрепляются на лабораторных занятиях и используются при выполнении самостоятельного исследования в рамках индивидуальной курсовой работы.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью освоения учебной дисциплины является овладение будущими специалистами теоретическими знаниями и практическими навыками работы с моделями машинного обучения:

- применение методов машинного обучения для решения прикладных задач обработки данных
- использование современных программных библиотек для построения и анализа моделей машинного обучения

Освоение основных моделей и методов машинного обучения и анализа данных необходимо для проведения самостоятельных исследований, разработки теоретических моделей в новой предметной области, применения современных математических теорий для решения прикладных задач обработки данных.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина «Анализ данных и машинное обучение: часть 1» относится к общенаучному циклу М1 образовательной программы и 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 «Введение в машинное обучение». В свою очередь, дисциплина является предшествующей и необходимой для изучения следующих курсов профессионального цикла подготовки:

- Анализ данных и машинное обучение: часть 2;
- 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
УКЦ-1 [1] – Способен решать исследовательские, научно-технические и производственные задачи в условиях неопределенности, в том числе выстраивать деловую коммуникацию и организовывать работу команды с использованием цифровых ресурсов и технологий в цифровой среде	З-УКЦ-1 [1] – Знать современные цифровые технологии, используемые для выстраивания деловой коммуникации и организации индивидуальной и командной работы У-УКЦ-1 [1] – Уметь подбирать наиболее релевантные цифровые решения для достижения поставленных целей и задач, в том числе в условиях неопределенности В-УКЦ-1 [1] – Владеть навыками решения исследовательских, научно-технических и производственных задач с использованием цифровых технологий

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 <i>Основание:</i> Профессиональный стандарт: 32.002	З-ПК-8.1[1] - знать методы прикладной математики, используемые при построении математических моделей киберфизических систем; У-ПК-8.1[1] - уметь обоснованно выбирать и применять методы

			<p>прикладной математики при разработке математического обеспечения киберфизических систем;</p> <p>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. ;</p> <p>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 прикладные результаты самостоятельно и в составе научного коллектива. ;</p> <p>В-ПК-1[1] - Владеть навыками выбора и</p>

			использования математических средств научных исследований, методами анализа и синтеза научной информации.
анализ и математическое моделирование физических процессов	системы ядерно-энергетического комплекса	ПК-2 [1] - способен к разработке и внедрению наукоемкого программного обеспечения, способствующего решению передовых задач науки и техники на основе современных математических методов и алгоритмов <i>Основание:</i> Профессиональный стандарт: 24.057	3-ПК-2[1] - Знать текущее положение современных научных достижений, современные математические методы и алгоритмы для разработки наукоемкого программного обеспечения. ; У-ПК-2[1] - Уметь применять современные математические методы и алгоритмы для разработки наукоемкого программного обеспечения.; В-ПК-2[1] - Владеть навыками разработки и внедрения наукоемкого программного обеспечения.
анализ и математическое моделирование физических процессов	системы ядерно-энергетического комплекса	ПК-3 [1] - способен развивать инновационный потенциал новых научных и научно-технологических разработок <i>Основание:</i> Профессиональный стандарт: 24.078	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

			<p>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 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; 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</p>
--	--	--	--

	<i>2 Семестр</i>						
1	Методы классификации данных	1-7	7/7/7		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-2, У-ПК-2, В-ПК-2, 3-ПК-3, У-ПК-3, В-ПК-3, 3-ПК-4, У-ПК-4, В-ПК-4, 3-УК-1, У-УК-1, В-УК-1, 3-УКЦ-1, У-УКЦ-1, В-УКЦ-1
2	Непараметрические методы и обучение ансамблей	8-15	8/8/8		25	КИ-15	3-ПК-8.1, У-ПК-8.1, В-ПК-8.1, 3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-2, У-ПК-2, В-ПК-2, 3-ПК-3, У-ПК-3, В-ПК-3, 3-ПК-4, У-ПК-4, В-ПК-4, 3-УК-1, У-УК-1, В-УК-1, 3-УКЦ-1, У-УКЦ-1, В-УКЦ-1
	<i>Итого за 2 Семестр</i>		15/15/15		50		
	Контрольные мероприятия за 2 Семестр				50	Э	3-ПК-1, У-ПК-1, В-ПК-1

* – 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>2 Семестр</i>	15	15	15
1-7	Методы классификации данных	7	7	7
1 - 3	Байесовская классификация Порождающие и разделяющие модели классификации. Бинарная функция потерь. Байесовское решающее правило. Дискриминантные функции. Нормальный байесовский классификатор. Расстояние Махаланобиса. Наивный байесовский классификатор. Регуляризованный байесовский классификатор. Байесовская классификация бинарных и мультиномиальных признаков. Пример применения к задаче категоризации документов.	Всего аудиторных часов		
		3	3	3
		Онлайн		
		0	0	0
4 - 5	Непараметрические методы классификации Непараметрический байесовский классификатор. Наивная оценка плотности распределения. Ядерная оценка плотности распределения. Виды ядерных функций. Парзеновское окно. Показатель точности оценки плотности. Разложение «bias-variance» ошибки оценивания. Подходы к выбору ширины парзеновского окна. Метод Сильвермана. Оценивание многомерных плотностей распределения. Многомерные ядерные функции.	Всего аудиторных часов		
		2	2	2
		Онлайн		
		0	0	0
6 - 7	Логистическая регрессия Дискриминативный подход к классификации. Функции связи в задаче классификации. Виды функций связи, их свойства. Логит и пробит-функции связи. Логистическая функция. Логистическая модель апостериорных вероятностей классов. Отношение вероятностей (odds ratio). Отношение вероятностей для нормального байесовского классификатора. Связь нормального байесовского классификатора и логистической регрессии. Постановка задачи обучения бинарной логистической регрессии. Критерий обучения и метод обучения. Логистическая функция потерь. Регуляризованная логистическая регрессия Понятие softmax-функции. Softmax-функция для двух переменных. Связь с логистической функцией. Softmax-модель апостериорных вероятностей классов. Референсный класс. Постановка задачи обучения многоклассовой логистической регрессии. Критерий обучения и метод обучения.	Всего аудиторных часов		
		2	2	2
		Онлайн		
		0	0	0
8-15	Непараметрические методы и обучение ансамблей	8	8	8

8 - 10	Непараметрические регрессионные модели Непараметрическое машинное обучение с учителем и без учителя. Параметрические и непараметрические регрессионные модели. Наивная регрессия. Использование ядерных функций в задачах регрессии. Ядерная регрессия Надарая-Уотсона. KNN- регрессия. Регрессия и линейное сглаживание. Свойства сглаживающих фильтров. Локальная регрессия. LOESS. Разложение «bias-variance» ошибки непараметрической регрессионной модели. Подходы к выбору ширины окна Надарая-Уотсона. Многомерный метод LOESS.	Всего аудиторных часов		
		3	3	3
		Онлайн		
		0	0	0
11 - 12	Деревья решений Понятие дерева решений. Типы и постановки решаемых задач. Типы узлов дерева. Обучение дерева решений. Критерий оптимальности. Примеры. Настройка гиперпараметров процедуры обучения. Переобучение и «обрезка» дерева.	Всего аудиторных часов		
		2	2	2
		Онлайн		
		0	0	0
13 - 15	Обучение ансамблей Подходы к построению ансамблей моделей машинного обучения. Бэггинг и бустинг. Бутстреп. Ансамбли деревьев решений. Усиленный ансамбль деревьев. Кросс-валидация усиленного ансамбля. Случайный лес. Базовый алгоритм бустинга. Принципы и типы адаптивного бустинга. AdaBoost. Пошаговое аддитивное обучение. Градиентный бустинг моделей машинного обучения. Регуляризованный градиентный бустинг. Стохастический градиентный бустинг. XGBoost.	Всего аудиторных часов		
		3	3	3
		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

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

Недели	Темы занятий / Содержание
	<i>2 Семестр</i>
1 - 2	Байесовская классификация Байесовская классификация.
3 - 4	Непараметрические методы классификации Непараметрические методы классификации.
5 - 6	Логистическая регрессия Логистическая регрессия.

9 - 10	Непараметрические регрессионные модели Непараметрические регрессионные модели.
11 - 12	Деревья решений. Деревья решений.
13 - 15	Обучение ансамблей. Обучение ансамблей.

ТЕМЫ СЕМИНАРОВ

Недели	Темы занятий / Содержание
	<i>2 Семестр</i>
1 - 2	Байесовская классификация Байесовская классификация.
3 - 4	Непараметрические методы классификации Непараметрические методы классификации.
5 - 6	Логистическая регрессия Логистическая регрессия.
9 - 10	Непараметрические регрессионные модели Непараметрические регрессионные модели.
11 - 12	Деревья решений Деревья решений.
13 - 15	Обучение ансамблей Обучение ансамблей.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

1. Лекционные занятия:

- a. комплект электронных презентаций/слайдов,
- b. аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук)

2. Практические и лабораторные занятия:

- a. компьютерный класс,
- b. презентационная техника (проектор, экран, компьютер/ноутбук),
- c. 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 (КП 1)
ПК-1	З-ПК-1	Э, КИ-8, КИ-15
	У-ПК-1	Э, КИ-8, КИ-15
	В-ПК-1	Э, КИ-8, КИ-15
ПК-2	З-ПК-2	КИ-8, КИ-15

	У-ПК-2	КИ-8, КИ-15
	В-ПК-2	КИ-8, КИ-15
ПК-3	З-ПК-3	КИ-8, КИ-15
	У-ПК-3	КИ-8, КИ-15
	В-ПК-3	КИ-8, КИ-15
ПК-4	З-ПК-4	КИ-8, КИ-15
	У-ПК-4	КИ-8, КИ-15
	В-ПК-4	КИ-8, КИ-15
ПК-8.1	З-ПК-8.1	КИ-8, КИ-15
	У-ПК-8.1	КИ-8, КИ-15
	В-ПК-8.1	КИ-8, КИ-15
УК-1	З-УК-1	КИ-8, КИ-15
	У-УК-1	КИ-8, КИ-15
	В-УК-1	КИ-8, КИ-15
УКЦ-1	З-УКЦ-1	КИ-8, КИ-15
	У-УКЦ-1	КИ-8, КИ-15
	В-УКЦ-1	КИ-8, КИ-15

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	4 – «хорошо»		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 77 Машинное обучение с участием человека : , Монарх Р., Москва: ДМК Пресс, 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/>

<http://library.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 заданий), 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.