

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

ОДОБРЕНО УМС ИИКС

Протокол № 8/1/2024

от 28.08.2024 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ

Направление подготовки
(специальность)

[1] 10.04.01 Информационная безопасность

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, экз./зач./КР/КП
2	2	72	30	0	30		12	0	З
3	3	108	32	0	32		8	0	Э
Итого	5	180	62	0	62	0	20	0	

АННОТАЦИЯ

В курсе рассматриваются следующие темы:

- основы интеллектуального анализа данных;
- основы методов машинного обучения;
- методы классификации, кластеризации и восстановления регрессии;
- основы методов глубокого обучения, включая искусственные нейронные сети.

В рамках лабораторного практикума студенты получают навыки программирования.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Цель дисциплины – освоение комплекса современных методов интеллектуального анализа данных и процессов, включая методы, основанные на машинном обучении.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Полученные в результате освоения учебной дисциплины знания, умения, навыки используются в процессе дипломного проектирования.

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные и(или) общепрофессиональные компетенции:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
--------------------------------	--

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
проектный			
разработка проектных решений по обеспечению безопасности данных с применением криптографических методов	информационные ресурсы	ПК-4.1 [1] - Способен разрабатывать проектные решения по обеспечению безопасности данных с применением криптографических методов <i>Основание:</i> Профессиональный стандарт: 06.032	З-ПК-4.1[1] - Знать: методы обеспечения безопасности данных с применением криптографических методов; У-ПК-4.1[1] - Уметь: разрабатывать проектные решения по обеспечению безопасности данных с применением

			криптографических методов; В-ПК-4.1[1] - Владеть: навыками разработки проектных решений по обеспечению безопасности данных с применением криптографических методов
разработка проектных решений по обеспечению безопасности данных с применением криптографических методов	информационные ресурсы	ПК-1 [1] - Способен принимать участие в разработке систем обеспечения ИБ или информационно-аналитических систем безопасности <i>Основание:</i> Профессиональный стандарт: 06.032	З-ПК-1[1] - Знать: модели угроз нсд к сетям электросвязи; методики оценки уязвимостей сетей электросвязи с точки зрения возможности нсд к ним; нормативные правовые акты в области связи, информатизации и защиты информации; виды политик безопасности компьютерных систем и сетей; возможности используемых и планируемых к использованию средств защиты информации; особенности защиты информации в автоматизированных системах управления технологическими процессами; критерии оценки эффективности и надежности средств защиты информации программного обеспечения автоматизированных систем; основные характеристики технических средств защиты информации от утечек по техническим каналам; нормативные правовые акты, методические документы, национальные

			стандарты в области защиты информации ограниченного доступа и аттестации объектов информатизации на соответствие требованиям по защите информации; технические каналы утечки информации. ; У-ПК-1[1] - Уметь: выявлять и оценивать угрозы нсд к сетям электросвязи; анализировать компьютерную систему с целью определения необходимого уровня защищенности и доверия; классифицировать защищаемую информацию по видам тайны и степеням конфиденциальности; выбирать меры защиты информации, подлежащие реализации в системе защиты информации автоматизированной системы; проводить анализ угроз безопасности информации на объекте информатизации; проводить предпроектное обследование объекта информатизации. ; В-ПК-1[1] - Владеть: основами проведения технических работ при аттестации сссэ с учетом требований по защите информации; определением угроз безопасности информации, реализация которых может привести к нарушению
--	--	--	---

			безопасности информации в компьютерной системе и сети; основами разработки модели угроз безопасности информации и модели нарушителя в автоматизированных системах; основами предпроектного обследования объекта информатизации; основами разработки аналитического обоснования необходимости создания системы защиты информации на объекте информатизации (модели угроз безопасности информации).
--	--	--	---

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>2 Семестр</i>						
1	Первый раздел	1-8	20/0/20		25	КИ-8	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1
2	Второй раздел	9-15	10/0/10		25	КИ-15	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1
	<i>Итого за 2 Семестр</i>		30/0/30		50		
	Контрольные мероприятия за 2				50	3	3-ПК-1, У-ПК-1,

	Семестр						В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1
	<i>3 Семестр</i>						
1	Первый раздел	1-8	16/0/16		25	КИ-8	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1
2	Второй раздел	9-15	16/0/16		25	КИ-15	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1
	<i>Итого за 3 Семестр</i>		32/0/32		50		
	Контрольные мероприятия за 3 Семестр				50	Э	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-4.1, У-ПК-4.1, В-ПК-4.1

* – 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
З	Зачет
Э	Экзамен

КАЛЕНДАРНЫЙ ПЛАН

Недели	Темы занятий / Содержание	Лек., час.	Пр./сем., час.	Лаб., час.
	<i>2 Семестр</i>	30	0	30
1-8	Первый раздел	20	0	20
1 - 2	Предмет анализа данных. Основные методы анализа данных. Отличительные особенности интеллектуального анализа данных. Основная терминология. Типы данных. Типы признаков. Приложения интеллектуального анализа данных. Жизненный цикл процесса анализа данных.	Всего аудиторных часов		
		5	0	5
		Онлайн		
		0	0	0
3 - 4	Предобработка и постобработка данных. Задачи, возникающие в связи с анализом данных. Взаимосвязь интеллектуального анализа данных с	Всего аудиторных часов		
		5	0	5
		Онлайн		

	другими видами обработки данных. Способы хранения данных. Инструментарий анализа данных. Основные задачи предобработки данных. Очистка данных. Интеграция данных. Редукция данных. Дискретное вейвлет-преобразование. Вейвлеты Хаара. Визуализация данных.	0	0	0
5 - 6	Поиск шаблонов и ассоциативных правил. Постановка задачи поиска часто встречающихся шаблонов и ассоциативных правил. Алгоритм Apriori. Генерация ассоциативных правил из часто встречающихся подмножеств. Алгоритм FP-Growth.	Всего аудиторных часов		
		5	0	5
		Онлайн		
		0	0	0
7 - 8	Введение в машинное обучение. Формальная постановка задачи машинного обучения. Задание объектов в машинном обучении. Задание ответов в машинном обучении. Этапы обучения и применения модели. Функционалы качества. Сведение задачи обучения к задаче оптимизации. Эмпирические оценки обобщающей способности. Способы проверки обобщающей способности на практике. Переобучение.	Всего аудиторных часов		
		5	0	5
		Онлайн		
		0	0	0
9-15	Второй раздел	10	0	10
9 - 10	Классификация. Формальная постановка задачи классификации. Логические методы классификации. Бинарные решающие деревья. Алгоритм построения решающего дерева по обучающей выборке. Варианты критериев ветвления. Индекс Джини. Достоинства и недостатки решающих деревьев. Проблемы, связанные с «жадным» алгоритмом ID3.	Всего аудиторных часов		
		3	0	3
		Онлайн		
		0	0	0
11 - 12	Оценка качества классификаторов. Метрики для оценки качества классификаторов. ROC-кривая. Композиции алгоритмов классификации. Простое голосование классификаторов.	Всего аудиторных часов		
		3	0	3
		Онлайн		
		0	0	0
13 - 14	Кластеризация. Формальная постановка задачи кластеризации. Расстояния между объектами в признаковом пространстве. Типы кластерных структур. Методы кластеризации. Метод k средних. Агломеративная иерархическая кластеризация. Формула Ланса – Уильямса. Визуализация кластерной структуры. Основные свойства иерархической кластеризации. Плотностные методы кластеризации. Алгоритм DBSCAN.	Всего аудиторных часов		
		2	0	2
		Онлайн		
		0	0	0
15	Обнаружение аномалий. Аномалии: определение и пример. Методы обнаружения аномалий. Методы, основанные на измерении близости объектов. Метод вложенных циклов. Сеточный метод. Кластерные методы. Примеры применения кластерных методов обнаружения аномалий в задачах информационной безопасности.	Всего аудиторных часов		
		2	0	2
		Онлайн		
		0	0	0
	<i>3 Семестр</i>	32	0	32
1-8	Первый раздел	16	0	16
1 - 2	Метрические методы классификации и восстановления регрессии.	Всего аудиторных часов		
		4	0	4

	Формальные постановки задач классификации и восстановления регрессии. Обобщенный метрический классификатор. Метод k ближайших соседей. Метод окна Парзена. Метод потенциальных функций. Метрические методы в задачах восстановления регрессии. Непараметрическая регрессия. Формула ядерного сглаживания Надарая – Ватсона. Часто используемые ядра. Выбор ядра и ширины окна. Проблема выбросов. Локально взвешенное сглаживание. Алгоритм LOWESS. Вероятностные (байесовские) методы классификации. Теорема Байеса. Наивный байесовский классификатор. Примеры применения. Метод коррекции Лапласа для случая отсутствия объектов с требуемыми значениями атрибутов.	Онлайн		
		0	0	0
3 - 4	Линейные классификаторы. Определение и формальное описание линейного классификатора. Обучение линейного классификатора. Непрерывные аппроксимации пороговой функции потерь. Обучение регрессии. Градиент функции. Градиентный метод численной минимизации. Метод стохастического градиента. Обоснование оценки функционала в методе стохастического градиента. Метод стохастического усредненного градиента. Достоинства и недостатки метода стохастического градиента. Эвристики: варианты инициализации весов, варианты порядка предъявления объектов. Проблема мультиколлинеарности. Проявления переобучения. Регуляризация.	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0
5 - 6	Логистическая регрессия. Обоснование логарифмической функции потерь. Оптимизация параметров логистической регрессии. Скоринговые карты. Регуляризованная логистическая регрессия. Композиции алгоритмов машинного обучения. Простое голосование классификаторов. Бэггинг и метод случайных подпространств. Метод случайного леса. Градиентный бустинг. Параметрическая аппроксимация градиентного шага. Стохастический градиентный бустинг. Алгоритм AdaBoost. Частные случаи при различных функциях потерь. Градиентный бустинг над деревьями.	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0
7 - 8	Логистическая регрессия. Задача частичного обучения. Отличия от задач классификации и кластеризации. Метод самообучения. Метод совместного обучения. Метод обучения композиции алгоритмов. Методы частичного обучения на основе кластеризации: графовый подход, метод Ланса – Уильямса, метод k средних.	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0
9-15	Второй раздел	16	0	16
9 - 10	Метод опорных векторов. Постановка задачи для метода опорных векторов. Первый подход – аппроксимация и регуляризация эмпирического риска. Второй подход – оптимальная разделяющая гиперплоскость. Переход к линейно неразделимой	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0

	выборке. Условия Каруша – Куна – Таккера. Двойственная задача и её нелинейное обобщение для метода опорных векторов. Ядра для нелинейного обобщения метода опорных векторов. Примеры ядер. Классификация с разными ядрами.			
11 - 12	Введение в искусственные нейронные сети. Линейная модель нейрона Мак Каллока – Питтса. Часто используемые функции активации нейронов. Нейронная реализация логических функций. Приближение произвольной функции нейронной сетью.	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0
13 - 14	Обучение искусственных нейронных сетей. Многослойная нейронная сеть. Алгоритм стохастического градиента в применении к нейронным сетям. Задача дифференцирования суперпозиции функций. Быстрое вычисление градиента. Алгоритм обратного распространения ошибки. Эвристики для обучения нейронных сетей. Ускорение сходимости. Динамическое наращивание сети. Прореживание сети (метод OBD). Увеличение числа уровней.	Всего аудиторных часов		
		4	0	4
		Онлайн		
		0	0	0
15	Введение в глубокое обучение. Свёрточные нейронные сети. Архитектура свёрточной сети. Отличия от полносвязной нейронной сети. Слой свёртки. Слой ReLU. Слой субдискретизации. Преимущества и недостатки свёрточных нейронных сетей. Применение свёрточных нейронных сетей в фотофоре́нзике.	Всего аудиторных часов		
		4	0	4
		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>
	Л/Р 1 Предобработка и постобработка данных.
	Л/Р 2 Поиск шаблонов и ассоциативных правил.
	Л/Р 3 Оценка качества классификаторов.
	Л/Р 4

	Обнаружение аномалий.
	3 Семестр
	Л/Р 1 Метрические методы классификации и восстановления регрессии.
	Л/Р 2 Линейные классификаторы.
	Л/Р 3 Метод опорных векторов.
	Л/Р 4 Обучение искусственных нейронных сетей.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Образовательные технологии сочетают в себе совокупность методов и средств для реализации определенного содержания обучения и воспитания в рамках дисциплины, включают решение дидактических и воспитательных задач, формируя основные понятия дисциплины, технологии проведения занятий, усвоения новых знаний, технологии повторения и контроля материала, самостоятельной работы.

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)	Аттестационное мероприятие (КП 2)
ПК-1	З-ПК-1	З, КИ-8, КИ-15	Э, КИ-8, КИ-15
	У-ПК-1	З, КИ-8, КИ-15	Э, КИ-8, КИ-15
	В-ПК-1	З, КИ-8, КИ-15	Э, КИ-8, КИ-15
ПК-4.1	З-ПК-4.1	З, КИ-8, КИ-15	Э, КИ-8, КИ-15
	У-ПК-4.1	З, КИ-8, КИ-15	Э, КИ-8, КИ-15
	В-ПК-4.1	З, КИ-8, КИ-15	Э, 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	Оценка ECTS	Требования к уровню освоению учебной дисциплины
--------------	------------------------------	-------------	---

90-100	5 – «отлично»	A	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы.
85-89	4 – «хорошо»	B	Оценка «хорошо» выставляется студенту, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос.
75-84		C	
70-74		D	
65-69	3 – «удовлетворительно»	E	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала.
60-64			
Ниже 60	2 – «неудовлетворительно»	F	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

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

Студенты должны своевременно спланировать учебное время для поэтапного и системного изучения данной учебной дисциплины в соответствии с планом лекций и семинарских занятий, графиком контроля знаний.

Успешное освоение дисциплины требует от студентов посещения лекций, активной работы во время семинарских занятий, выполнения всех домашних заданий, ознакомления с базовыми учебниками, основной и дополнительной литературой, а также предполагает творческое участие студента путем планомерной, повседневной работы.

Изучение дисциплины следует начинать с проработки учебной программы, особое внимание, уделяя целям и задачам, структуре и содержанию курса.

Во время лекций рекомендуется писать конспект. Запись лекции – одна из форм активной самостоятельной работы студентов, требующая навыков и умения кратко, схематично, последовательно и логично фиксировать основные положения, выводы, обобщения, формулировки.

При необходимости в конце лекции преподаватель оставляет время для того, чтобы студенты имели возможность задать вопросы по изучаемому материалу.

Лекции нацелены на освещение основополагающих положений теории алгоритмов и теории функций алгебры логики, наиболее трудных вопросов, как правило, связанных с доказательством необходимых утверждений и теорем, призваны способствовать формированию навыков работы с научной литературой. Предполагается также, что студенты приходят на лекции, предварительно проработав соответствующий учебный материал по источникам, рекомендуемым программой.

Конспект лекций для закрепления полученных знаний необходимо просмотреть сразу после занятий. Хорошо отметить материал конспекта лекций, который вызывает затруднения для понимания. Можно попытаться найти ответы на затруднительные вопросы, используя рекомендуемую литературу. Если самостоятельно не удалось разобраться в материале, рекомендуется сформулировать вопросы и обратиться за помощью к преподавателю на консультации или ближайшей лекции.

В процессе изучения учебной дисциплины необходимо обратить внимание на самоконтроль. Требуется регулярно отводить время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам, а также для выполнения домашних заданий, которые выдаются после каждого семинара.

Систематическая индивидуальная работа, постоянная активность на занятиях, готовность ставить и обсуждать актуальные проблемы курса – залог успешной работы и положительной оценки.

10. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

Учебный курс строится на интегративной основе и включает в себя как теоретические знания, так и практические навыки, получаемые студентами в ходе лекций, аудиторных практических занятий, лабораторных и самостоятельных занятий.

Данная дисциплина выполняет функции теоретической и практической подготовки студентов. Содержание дисциплины распределяется между лекционной и практической

частями на основе принципа дополняемости: практические занятия, как правило, не дублируют лекции и посвящены рассмотрению практических примеров и конкретизации материала, введенного на лекции. В лекционном курсе главное место отводится общетеоретическим проблемам.

Содержание учебного курса, его объем и характер обуславливают необходимость оптимизации учебного процесса в плане отбора материала обучения и методики его организации, а также контроля текущей учебной работы. В связи с этим возрастает значимость и изменяется статус внеаудиторной (самостоятельной) работы, которая становится полноценным и обязательным видом учебно-познавательной деятельности студентов. При изучении курса самостоятельная работа включает:

- самостоятельное ознакомление студентов с теоретическим материалом, представленным в отечественных и зарубежных научно-практических публикациях;

- самостоятельное изучение тем учебной программы, достаточно хорошо обеспеченных литературой и сравнительно несложных для понимания;

- подготовку к практическим занятиям по тем разделам, которые не дублируют темы лекционной части, а потому предполагают самостоятельную проработку материала учебных пособий.

Со стороны преподавателя должен быть установлен контакт со студентами, и они должны быть информированы о порядке прохождения курса, его особенностях, учебно-методическом обеспечении по данной дисциплине. Преподаватель дает методические рекомендации обучаемым по самостоятельному изучению проблем, характеризуя пути и средства достижения поставленных перед ними задач, высказывает советы и рекомендации по изучению учебной литературы, самостоятельной работе и работе на семинарских занятиях.

Автор(ы):

Запечников Сергей Владимирович, д.т.н., доцент