Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»

ИНСТИТУТ ФИНАНСОВЫХ ТЕХНОЛОГИЙ И ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ

КАФЕДРА ФИНАНСОВОГО МОНИТОРИНГА

ОДОБРЕНО УМС ИФТЭБ

Протокол № 545-6

от 28.08.2025 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

МАШИННОЕ ОБУЧЕНИЕ

Направление подготовки (специальность) — [1] 10.05.04 Информационно-аналитические системы безопасности

Семестр	**Трудоемкость, кред.**	**Общий объем курса, час.**	**Лекции, час.**	**Практич. занятия, час.**	**Лаборат. работы, час.**	**В форме практической подготовки, час.**	**СРС, час.**	**КСР, час.**	**Форма(ы) контроля, экз./зач./КР/КП**
7	2-3	72-108	16	16	24		16-52	0	З
8	4	144	15	15	22		38-47	0	Э
Итого	6-7	216-252	31	31	46	22	54-99	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

Целями освоения дисциплины является изучение математических основ и технологий машинного обучения, получение представления об основных задачах машинного обучения, построения и обучения нейронных сетей.

Задачи курса:
- сформировать теоретические знания по основам машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования;
- выработать умения по практическому применению методов машинного обучения при решении прикладных задач в различных областях;
- выработать умения и навыки использования библиотек языка Python для разработки систем машинного обучения.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Для изучения курса необходимы базовые знания информатики, линейной алгебры, математического анализа, теории вероятностей, математической статистики, методов оптимизации.

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**
ОПК-2.4 [1] – Способен разрабатывать и применять автоматизированные технологии обработки больших информационных потоков (массивов) финансовой и/или экономической информации в режиме реального времени	З-ОПК-2.4 [1] – знать основные существующие автоматизированные технологии обработки больших информационных потоков (массивов) финансовой и/или экономической информации в режиме реального времени У-ОПК-2.4 [1] – уметь разрабатывать и применять автоматизированные технологии обработки больших информационных потоков (массивов) финансовой и/или экономической информации в режиме реального времени В-ОПК-2.4 [1] – владеть навыками разработки и применения автоматизированных технологий обработки

	больших информационных потоков (массивов) финансовой и/или экономической информации в режиме реального времени

4. ВОСПИТАТЕЛЬНЫЙ ПОТЕНЦИАЛ ДИСЦИПЛИНЫ

Направления/цели воспитания	**Задачи воспитания (код)**
Профессиональное воспитание	Создание условий, обеспечивающих, формирование культуры финансовой безопасности (В44)

5. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

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

№ п.п	**Наименование раздела учебной дисциплины**	**Недели**	**Лекции/ Практ. (семинары)/ Лабораторные работы, час.**	**Обязат. текущий контроль (форма*, неделя)**	**Максимальный балл за раздел****	**Аттестация раздела (форма*, неделя)**	**Индикаторы освоения компетенции**
	7 Семестр						
1	Основы машинного обучения и инструменты анализа данных	1-8	8/8/12	Отч-8 (25)	25	КИ-8	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4
2	Классификация данных: алгоритмы, снижение размерности и оценка качества	9-16	8/8/12	Отч-15 (25)	25	КИ-16	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4
	Итого за 7 Семестр		16/16/24		50		
	Контрольные мероприятия за 7 Семестр				50	З	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4
	8 Семестр						
1	Ансамблевые методы классификации и методы кластеризации	1-8	8/8/12	Отч-8 (25)	25	КИ-8	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4
2	Продвинутые методы кластеризации и основы нейронных сетей	9-15	7/7/10	Отч-14 (25)	25	КИ-15	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4
	Итого за 8 Семестр		15/15/22		50		

	Контрольные мероприятия за 8 Семестр				50	Э	З-ОПК-2.4, У-ОПК-2.4, В-ОПК-2.4

* – 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**
Отч	Отчет
КИ	Контроль по итогам
З	Зачет
Э	Экзамен

КАЛЕНДАРНЫЙ ПЛАН

Недели	**Темы занятий / Содержание**	**Лек., час.**	**Пр./сем., час.**	**Лаб., час.**
	7 Семестр	16	16	24
1-8	**Основы машинного обучения и инструменты анализа данных**	8	8	12
1 - 2	**Тема 1. Введение в машинное обучение. Язык программирования Python.** 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		
		2	2	3
		Онлайн		
		0	0	0
3 - 4	**Тема 2. Pandas - программная библиотека на языке Python.** Изучение основ библиотеки Pandas для работы с большими данными.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
5 - 6	**Тема 3. Корреляционный анализ. Анализ тесноты взаимосвязи для количественных данных.** Корреляционный анализ. Анализ тесноты взаимосвязи для количественных данных.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
7 - 8	**Тема 4. Корреляционный анализ. Анализ тесноты взаимосвязи для качественных данных. Часть 1.** Корреляционный анализ. Анализ тесноты взаимосвязи для качественных данных.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
9-16	**Классификация данных: алгоритмы, снижение размерности и оценка качества**	8	8	12
9 - 10	**Тема 5. Корреляционный анализ. Анализ тесноты взаимосвязи для качественных данных. Часть 2.** Корреляционный анализ. Анализ тесноты взаимосвязи для качественных данных.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0

11 - 12	**Тема 6. Классификация данных с помощью древовидных структур.** Классификация. Формальная постановка задачи классификации. Логические методы классификации. Бинарные решающие деревья. Деревья решений. Алгоритм построения решающего дерева по обучающей выборке. Варианты критериев ветвления.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
13 - 14	**Тема 7. Классификация. Метод главных компонент. Метод опорных векторов** Снижение размерности данных и выявление наиболее информативных признаков в них с помощью метода главных компонент . Классификация данных с помощью метода опорных векторов (SVM).	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
15 - 16	**Тема 8. Оценка качества классификаторов. Метрики.** Оценка качества классификаторов. Метрики для оценки качества классификаторов. ROC-кривая. Композиции алгоритмов классификации. Простое голосование классификаторов.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
	8 Семестр	15	15	22
1-8	**Ансамблевые методы классификации и методы кластеризации**	8	8	12
1 - 2	**Тема 1. Классификация данных с помощью с помощью метода случайного леса.** Алгоритм машинного обучения Random Forest. Подготовка выборки. Случайный лес.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
3 - 4	**Тема 2. Классификация данных с помощью бустинга.** Градиентный бустинг. Параметрическая аппроксимация градиентного шага. Стохастический градиентный бустинг. Алгоритм AdaBoost. Частные случаи при различных функциях потерь. Градиентный бустинг над деревьями.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
5 - 6	**Тема 3. Кластеризация данных. Метод k-средних.** Формальная постановка задачи кластеризации. Расстояния между объектами в признаковом пространстве. Типы кластерных структур. Методы кластеризации. Метод k средних.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
7 - 8	**Тема 4. Кластеризация данных. Агломеративная кластеризация.** Агломеративная иерархическая кластеризация. Формула Ланса – Уильямса. Визуализация кластерной структуры. Основные свойства иерархической кластеризации. Плотностные методы кластеризации. Алгоритм DBSCAN.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
9-15	**Продвинутые методы кластеризации и основы нейронных сетей**	7	7	10
9 - 10	**Тема 5. Кластеризация данных. Методы Spectral Clustering, Affinity Propagation.** Спектральная кластеризация. Метод распространения близости (affinity propagation). Модификации метода Affinity Propagation.	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0
11 - 12	**Тема 6. Введение в искусственные нейронные сети.** Линейная модель нейрона Мак Каллока – Питтса. Часто используемые функции активации нейронов. Нейронная реализация логических функций. Приближение	Всего аудиторных часов		
		2	2	3
		Онлайн		
		0	0	0

	произвольной функции нейронной сетью.			
13	**Тема 7. Классификации данных с помощью нейронных сетей.** Подготовка исходных данных. Кодирование выходных значений. Выбор размера сети. Выбор архитектуры сети. Алгоритм построения классификатора на основе нейронной сети.	Всего аудиторных часов		
		1	1	2
		Онлайн		
		0	0	0
14 - 15	**Тема 8. 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		
		2	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

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

Недели	**Темы занятий / Содержание**
	7 Семестр
1 - 2	**Лабораторная работа №1.** Pandas - изучение основ библиотеки для работы с большими данными.
3 - 5	**Лабораторная работа №2.** Корреляционный анализ - исследование реальных статистических данных на нормальное распределение с помощью методов Шапиро-Уилка и других, а также исследование на предмет взаимной корреляции.
6 - 8	**Лабораторная работа №3.** Хи-квадрат Пирсона - проверка нормальности распределения с помощью метода Хи-квадрат Пирсона больших статистических данных, гомоскедастичность и автокорреляция.
9 - 12	**Лабораторная работа №4.** Дерево принятия решений - изучение основ классификации данных с помощью древовидных структур.
13 - 15	**Лабораторная работа №5.** SVM - изучение основ классификации данных с помощью метода опорных векторов и расчет характеристик качества классификатора.
	8 Семестр

1 - 2	**Лабораторная работа №1.** Лес - изучение основ классификации данных с помощью уже изученных структур, их сравнительный анализ с результатами, полученными с помощью Random Forest - классификатора.
3 - 5	**Лабораторная работа №2.** Бустинг - изучение основ классификации данных с помощью бустинга: AdaBoost и Градиентного бустинга.
6 - 8	**Лабораторная работа №3.** Кластеризация, метод K-means - изучение основ кластеризации данных с помощью метода k-средних и агломеративных методов.
9 - 12	**Лабораторная работа №4.** Кластеризация, другие методы - изучение основ кластеризации данных с помощью методов DBSCAN, Spectral Clustering, Affinity Propagation.
13 - 15	**Лабораторная работа №5.** Нейронные сети - изучение основ классификации данных с помощью нейронных сетей в сравнении с изученными ранее классификаторами

6. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Используется чтение лекций по дисциплине с использованием компьютерного мультимедийного оборудования, предусматривающее разбор практических задач, моделирование типовых ситуаций, возникающих при решении задач анализа данных. На практике используются инструментальные средства (системы программирования, аналитические платформы, офисный пакет), позволяющие получить базовые навыки работы.

7. 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)**
ОПК-2.4	З-ОПК-2.4	З, КИ-8, КИ-16, Отч-8, Отч-15	Э, КИ-8, КИ-15, Отч-8, Отч-14
	У-ОПК-2.4	З, КИ-8, КИ-16, Отч-8, Отч-15	Э, КИ-8, КИ-15, Отч-8, Отч-14
	В-ОПК-2.4	З, КИ-8, КИ-16, Отч-8, Отч-15	Э, КИ-8, КИ-15, Отч-8, Отч-14

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:

<table>
<tr><th>Сумма баллов</th><th>Оценка по 4-х балльной шкале</th><th>Отметка о зачете</th><th>Оценка ECTS</th></tr>
<tr><td>90-100</td><td>5 – «отлично»</td><td rowspan="7">«зачтено»</td><td>A</td></tr>
<tr><td>85-89</td><td rowspan="3">4 – «хорошо»</td><td>B</td></tr>
<tr><td>75-84</td><td>C</td></tr>
<tr><td>70-74</td><td rowspan="2">D</td></tr>
<tr><td>65-69</td><td rowspan="2">3 – «удовлетворительно»</td></tr>
<tr><td>60-64</td><td>E</td></tr>
<tr><td>ниже 60</td><td>2 – «неудовлетворительно»</td><td>«не зачтено»</td><td>F</td></tr>
</table>

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.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ П 37 Python для сложных задач: наука о данных и машинное обучение. : , Плас Дж. Вандер, Санкт-Петербург: Питер, 2021

2. ЭИ Ю 68 Анализ больших наборов данных : , Ананд Р. , Юре Л. , Джеффри Д. У., Москва: ДМК Пресс, 2016

3. ЭИ Б 20 Машинное обучение и искусственный интеллект : учебное пособие для вузов, Баланов А. Н., Санкт-Петербург: Лань, 2024

4. ЭИ М 77 Машинное обучение с участием человека : , Монарх Р., Москва: ДМК Пресс, 2022

5. 004 А 51 Машинное обучение: новый искусственный интеллект : пер. с англ., Алпайдин Э., Москва: Альпина Паблишер, 2017

6. 004 С 36 Основы Data Science и Big Data. Python и наука о данных : , Силен Д., Мейсман А., Али М., Санкт-Петербург: Питер, 2020

7. ЭИ З-31 Основы интеллектуального анализа данных и машинного обучения : Конспект лекций. Учебное пособие, Запечников С.В., Москва: НИЯУ МИФИ, 2022

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. ЭИ Ю 68 Анализ больших наборов данных : , Ананд Р. , Юре Л. , Джеффри Д. У., Москва: ДМК Пресс, 2016

2. ЭИ М 63 Базовые методы анализа данных : учебник и практикум для вузов, Миркин Б. Г., Москва: Юрайт, 2024

3. ЭИ М 97 Вероятностное машинное обучение. Введение : учебное пособие, Мэрфи К. П., Москва: ДМК Пресс, 2022

4. ЭИ Б 20 Машинное обучение и искусственный интеллект : учебное пособие для вузов, Баланов А. Н., Санкт-Петербург: Лань, 2024

5. 004 А 51 Машинное обучение: новый искусственный интеллект : пер. с англ., Алпайдин Э., Москва: Альпина Паблишер, 2017

6. 004 С 36 Основы Data Science и Big Data. Python и наука о данных : , Силен Д., Мейсман А., Али М., Санкт-Петербург: Питер, 2020

7. ЭИ Б 20 Создание цифровых экосистем : учебное пособие для вузов, Баланов А. Н., Санкт-Петербург: Лань, 2024

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

Специальное программное обеспечение не требуется

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

1. Научная электронная библиотека «КиберЛенинка» (http://cyberleninka.ru)

2. Общероссийский математический портал Math-Net.Ru (http://www.mathnet.ru)

3. Образовательный портал GeekBrains (https://gb.ru)

4. Обучающие статьи о Computer Science и использование классических алгоритмов и структур данных в реше (https://tproger.ru/tag/algorithms/)

5. Электронная библиотека по техническим наукам (http://techlibrary.ru)

6. Информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллек (http://www.machinelearning.ru)

7. Pandas - библиотека Python для машинного обучения (https://pandas.pydata.org)

8. Обучение программированию на Python с помощью практических примеров (https://realpython.com)

https://online.mephi.ru/

http://library.mephi.ru/

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

10. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

В ходе освоения дисциплины студенты должны получить целостное представление о машинном обучении как о комплексе методов построения моделей по данным, сочетающем математическую формализацию задач и практическую реализацию алгоритмов. Изучение материала следует выстраивать от базовых понятий к сложным моделям, уделяя особое внимание взаимосвязи теоретических положений и их программной реализации.

Рекомендуется уделять значительное внимание языку Python как основному инструменту работы с данными и реализации алгоритмов машинного обучения. При выполнении практических заданий целесообразно активно использовать библиотеку Pandas для загрузки, очистки и преобразования данных: освоение типичных операций с DataFrame (фильтрация, группировка, агрегация, обработка пропусков) рассматривается как необходимая база для дальнейшей работы с моделями. Студентам рекомендуется закреплять навыки работы с Pandas на реальных наборах данных, обращая внимание на соответствие типов признаков и требования алгоритмов к формату входных данных.

При изучении методов анализа взаимосвязей между признаками рекомендуется различать типы данных (количественные и качественные) и подбирать соответствующие меры тесноты связи. Следует практиковаться в интерпретации коэффициентов корреляции и статистических показателей, сопоставляя их с визуальными представлениями (диаграммы рассеяния, тепловые карты корреляций). При этом важно понимать, что наличие корреляции не означает причинно-следственной связи, и учитывать контекст предметной области при формулировании выводов.

Освоение методов классификации рекомендуется сопровождать проработкой формальной постановки задачи, включая определение пространства признаков, целевой переменной и функционала качества. При построении моделей на основе деревьев решений, метода опорных векторов и ансамблевых подходов (случайный лес, бустинг) рекомендуется уделять внимание выбору гиперпараметров и оценке обобщающей способности модели. Для этого целесообразно применять кросс-валидацию и отдельные тестовые выборки, избегая оценки качества на тех же данных, на которых модель обучалась.

Изучение методов снижения размерности (в том числе метода главных компонент) рекомендуется рассматривать в контексте подготовки данных для моделей: выявление наиболее информативных признаков и устранение избыточности способствуют повышению устойчивости моделей и снижению риска переобучения. При работе с методами кластеризации (k-средних, иерархическая кластеризация, DBSCAN, спектральная кластеризация и др.) рекомендуется обращать внимание на выбор метрики расстояния, чувствительность алгоритмов к параметрам и интерпретацию полученных кластерных структур. Визуализация результатов кластеризации рассматривается как обязательный этап анализа.

При знакомстве с искусственными нейронными сетями рекомендуется сначала освоить базовые понятия (модель нейрона, функции активации, архитектура сети), а затем переходить к алгоритмам обучения, включая метод стохастического градиента и обратное распространение ошибки. При построении классификаторов на основе нейронных сетей рекомендуется уделять внимание подготовке данных (нормализация, кодирование категориальных признаков), выбору архитектуры и мониторингу процесса обучения (анализ кривых потерь и точности на обучающей и валидационной выборках). Следует учитывать, что обучение нейронных сетей требует итеративного подхода, и достижение приемлемого качества модели зачастую связано с подбором гиперпараметров и применением эвристик для ускорения сходимости и предотвращения переобучения.

Для закрепления теоретических знаний и развития практических навыков рекомендуется регулярно выполнять лабораторные работы и индивидуальные задания, документируя ход решения и обосновывая выбор методов и параметров моделей. Рекомендуется сохранять код в структурированном виде, сопровождая его комментариями и описанием этапов обработки данных и построения моделей. При возникновении трудностей целесообразно обращаться к официальной документации библиотек, научным статьям и учебным ресурсам, а также обсуждать проблемные моменты на консультациях с преподавателем.

Оценка результатов работы должна основываться не только на численных метриках качества, но и на понимании ограничений используемых методов, способности интерпретировать результаты и аргументировать выбор подхода для конкретной задачи. Рекомендуется развивать навык критического анализа получаемых решений, учитывая баланс между сложностью модели и ее интерпретируемостью, а также вычислительными затратами на обучение и применение.

11. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

Преподавателю рекомендуется выстраивать курс по принципу «от фундаментальных понятий к сложным моделям», обеспечивая четкую увязку теории и практики. Важно соблюдать баланс между математической формализацией задач (постановка, функционал качества, оптимизация) и прикладными навыками (работа с данными, реализация и настройка моделей). Для сопоставимости подходов целесообразно заранее подготовить единый набор учебных датасетов — на них студенты смогут отрабатывать разные методы и наглядно видеть различия в их работе.

В преподавании стоит делать акцент на типовых конвейерах обработки данных — от загрузки и проверки типов до обработки пропусков и выбросов, кодирования признаков и масштабирования. При разборе алгоритмов недостаточно показывать только реализацию: важно демонстрировать интерпретацию результатов с помощью визуализации (разбиения

пространства признаков, ROC-кривые, дендрограммы, графики потерь и точности) и отдельно подчеркивать, что корреляция не означает причинно-следственной связи.

Особое внимание следует уделять типичным трудностям студентов: непониманию разницы между обучающей, валидационной и тестовой выборками; ошибкам в предобработке (например, отсутствию масштабирования для SVM и нейросетей или некорректному кодированию категориальных признаков); неверной интерпретации метрик (в частности, использованию accuracy на несбалансированных данных без учета Precision/Recall). Эффективным приемом считается разбор таких ошибок на реальных примерах кода, а также регулярные консультации, где студенты могут показать промежуточные результаты и получить обратную связь.

Для формирования инженерного мышления рекомендуется сочетать лекции (для изложения теории и математических основ) с практическими и лабораторными занятиями (для отработки пайплайнов, настройки и интерпретации моделей), а также включать мини-проекты, охватывающие полный цикл работы — от постановки задачи до финального решения и выводов. Полезно использовать единые шаблоны отчётов по работам: они помогают студентам выработать навыки документирования ML-проектов и упрощают проверку.

Текущий контроль целесообразно строить на коротких тестах по ключевым понятиям, заданиях на интерпретацию графиков и мини-кейсах, где студент должен обосновать выбор метода и параметров.

Важно обеспечить студентам доступ к единой среде разработки и стандартным библиотекам для анализа данных и машинного обучения. При ограниченных вычислительных ресурсах рекомендуется использовать упрощённые датасеты и примеры, стабильно работающие на типовых конфигурациях. Оперативная обратная связь и возможность обсудить промежуточные результаты позволяют своевременно корректировать ход освоения курса и повышать качество подготовки обучающихся.

Автор(ы):

Радыгин Виктор Юрьевич