

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
МИНИСТЕРСТВО НАУКИ, ВЫСШЕГО ОБРАЗОВАНИЯ И ИННОВАЦИЙ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ

МОО ВО Кыргызско-Российский Славянский университет
имени первого Президента Российской Федерации Б.Н. Ельцина

УТВЕРЖДАЮ

и.о. декана ЕТФ

Комарцов Н.М.

Машинное обучение и анализ данных

4.11.24

рабочая программа дисциплины (модуля)

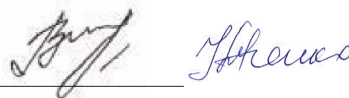
Закреплена за кафедрой	Информационных и вычислительных технологий		
Учебный план	g090404_24_12пи_рпис.plx Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия Магистерская программа "Разработка программно-информационных систем"		
Квалификация	магистр		
Форма обучения	очная		
Общая трудоемкость	4 ЗЕТ		
Часов по учебному плану	144	Виды контроля в семестрах: зачет с оценкой 2	
в том числе:			
аудиторные занятия	38		
самостоятельная работа	105		

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	2 (1.2)		Итого	
Неделя	18			
Вид занятий	УП	РП	УП	РП
Лекции	16	16	16	16
Практические	22	22	22	22
Контактная работа в период теоретического обучения	1	1	1	1
В том числе инт.	8	8	8	8
В том числе в форме практ.подготовки	22		22	
Итого ауд.	38	38	38	38
Контактная работа	39	39	39	39
Сам. работа	105	105	105	105
Итого	144	144	144	144

Программу составил(и):

д.т.н., профессор, Верзунов С.Н.; д.т.н., зав. каф., Лыченко Н.М.



Рабочая программа дисциплины

разработана в соответствии с ФГОС 3++:

Федеральный государственный образовательный стандарт высшего образования - магистратура по направлению подготовки 09.04.04 Программная инженерия (приказ Минобрнауки России от 19.09.2017 г. № 932)

составлена на основании учебного плана:

Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия

Магистерская программа "Разработка программно-информационных систем"

утвержденного учёным советом вуза от 22.10.24 протокол № 2

Рабочая программа одобрена на заседании кафедры

Протокол от 02.10.2024 г. № 2

Срок действия программы: 2025-2030 уч.г.

Зав. кафедрой Лыченко Н.М.



Визирование РПД для исполнения в очередном учебном году

Председатель УМС

9 сентября 2025 г. 

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2025-2026 учебном году на заседании кафедры
Информационных и вычислительных технологий

Протокол от 3 сентября 2025 г. № 1
Зав. кафедрой д.т.н., проф. Лыченко Н.М.



Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2026 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2026-2027 учебном году на заседании кафедры
Информационных и вычислительных технологий

Протокол от _____ 2026 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2027 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2027-2028 учебном году на заседании кафедры
Информационных и вычислительных технологий

Протокол от _____ 2027 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2028 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2028-2029 учебном году на заседании кафедры
Информационных и вычислительных технологий

Протокол от _____ 2028 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Цели освоения дисциплины (модуля)
1.2	Целью освоения дисциплины «Введение в машинное обучение» является формирование у студентов фундаментальных знаний, навыков и компетенций в области машинного обучения, позволяющих разрабатывать, анализировать и применять алгоритмы обработки данных для решения практических задач в программной инженерии.
1.3	Задачи курса
1.4	Ознакомить студентов с ключевыми концепциями машинного обучения, его методами и областями применения.
1.5	Научить студентов использовать популярные библиотеки Python (NumPy, Pandas, Matplotlib, Scikit-Learn) для предобработки, анализа и визуализации данных.
1.6	Изучить основные алгоритмы машинного обучения, включая линейную и логистическую регрессию, деревья решений, метод опорных векторов, случайные леса и алгоритмы кластеризации.
1.7	Дать представление о методах повышения качества моделей, таких как регуляризация, кросс-валидация, подбор гиперпараметров и ансамблирование (бустинг, случайные леса).
1.8	Рассмотреть принципы обработки естественного языка (NLP) и их применение в анализе текстовых данных.
1.9	Научить студентов применять полученные знания для решения реальных задач, в том числе предсказательного моделирования, анализа данных и создания рекомендательных систем.
1.10	Развить практические навыки работы с инструментами машинного обучения и подготовки данных для моделирования.
1.11	В результате освоения дисциплины студенты получают базовые знания и практические навыки, необходимые для работы в сфере Data Science, искусственного интеллекта и анализа данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:		Б1.В
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Для успешного освоения дисциплины «Введение в машинное обучение» студент должен обладать следующими знаниями и навыками:	
2.1.2	1. Математическая подготовка:	
2.1.3	Основы линейной алгебры (матрицы, векторы, операции с ними).	
2.1.4	Основы математического анализа (производные, интегралы, оптимизация).	
2.1.5	Основы теории вероятностей и статистики (среднее, дисперсия, распределения, корреляция).	
2.1.6	2. Программирование и алгоритмы:	
2.1.7	Базовые знания языка Python (переменные, функции, циклы, условия).	
2.1.8	Опыт работы с основными структурами данных (списки, словари, множества, кортежи).	
2.1.9	Понимание принципов алгоритмизации и написания программ.	
2.1.10	3. Работа с инструментами и библиотеками:	
2.1.11	Опыт работы в Jupyter Notebook или других средах разработки Python.	
2.1.12	Знание основ работы с библиотеками NumPy, Pandas, Matplotlib.	
2.1.13	4. Базовые знания в области обработки данных:	
2.1.14	Чтение, запись и обработка данных в различных форматах (CSV, JSON, Excel).	
2.1.15	Основы работы с базами данных и SQL (желательно).	
2.1.16	Навыки визуализации данных.	
2.1.17	Желательно, но не обязательно:	
2.1.18	Опыт работы с базами данных и SQL.	
2.1.19	Знание основ объектно-ориентированного программирования (ООП).	
2.1.20	Понимание принципов оптимизации и многопоточного программирования.	
2.1.21	Эти знания и навыки позволят студентам эффективно осваивать курс, понимать основные концепции машинного обучения и применять их на практике.	
2.1.22	Теория вероятностей и математическая статистика	
2.1.23	Математические основы программной инженерии	
2.1.24	Структурное программирование	
2.1.25	Вычислительная математика	
2.1.26	Алгоритмы и структуры данных	
2.1.27	Методы оптимизации	

2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	Освоение дисциплины «Введение в машинное обучение» является важным этапом подготовки студентов и служит основой для дальнейшего изучения ряда дисциплин и практик в области программной инженерии, анализа данных и искусственного интеллекта.
2.2.2	1. Дисциплины (модули), опирающиеся на данный курс:
2.2.3	Машинное обучение и глубокое обучение
2.2.4	(Углубленное изучение методов машинного обучения, нейросетей, алгоритмов глубокого обучения, работа с фреймворками TensorFlow и PyTorch.)
2.2.5	Анализ данных и Data Science
2.2.6	(Развитие компетенций по сбору, обработке и анализу данных, работа с большими данными.)
2.2.7	Компьютерное зрение
2.2.8	(Использование алгоритмов машинного обучения для обработки и анализа изображений, нейросетевые методы компьютерного зрения.)
2.2.9	Обработка естественного языка (NLP)
2.2.10	(Продвинутые методы обработки текстовых данных, работа с моделями на основе глубоких нейросетей.)
2.2.11	Разработка интеллектуальных систем
2.2.12	(Создание рекомендательных систем, чат-ботов, экспертных систем.)
2.2.13	Методы искусственного интеллекта
2.2.14	(Исследование различных подходов к созданию интеллектуальных систем, включая символические методы, алгоритмы обучения и адаптивные системы.)
2.2.15	2. Практики, требующие знаний по данному курсу:
2.2.16	Практика по анализу данных и моделированию
2.2.17	(Работа с реальными датасетами, применение методов машинного обучения, написание аналитических отчетов.)
2.2.18	Научно-исследовательская практика
2.2.19	(Разработка и тестирование алгоритмов машинного обучения, работа над научными статьями и исследовательскими проектами.)
2.2.20	Преддипломная практика
2.2.21	(Применение методов машинного обучения и анализа данных в реальных проектах.)
2.2.22	Проектная деятельность по программной инженерии
2.2.23	(Разработка программных решений с элементами машинного обучения, анализ данных для бизнес-приложений.)
2.2.24	Таким образом, данная дисциплина закладывает фундаментальные знания и навыки, необходимые для дальнейшего изучения искусственного интеллекта, анализа данных, компьютерного зрения, обработки текстов и разработки интеллектуальных систем.
2.2.25	Выполнение и защита выпускной квалификационной работы
2.2.26	Технологическая (проектно-технологическая) практика
2.2.27	Технологическая (проектно-технологическая) практика
2.2.28	Преддипломная практика 2
2.2.29	Преддипломная практика 1
2.2.30	Нейронные сети

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-4: Владение навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов

Знать:

Уровень 1	- методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов.
-----------	---

Уметь:

Уровень 1	- использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов.
-----------	--

Владеть:

Уровень 1	- методами и средствами создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов с помощью современных компьютерных
-----------	--

В результате освоения дисциплины обучающийся должен

3.1	Знать:
3.1.1	Основные принципы моделирования данных и методы предобработки данных.
3.1.2	Теоретические основы машинного обучения, включая методы обучения с учителем и без учителя.
3.1.3	Методы регрессии, классификации, кластеризации и принципы их работы.
3.1.4	Основы статистического анализа и способы оценки качества моделей.
3.1.5	Базовые алгоритмы и принципы работы нейросетевых моделей.
3.1.6	Методы визуализации данных для интерпретации результатов моделирования.
3.1.7	Принципы обработки естественного языка (NLP) и их применение.
3.1.8	Основы программирования на Python и работы с библиотеками NumPy, Pandas, Matplotlib, Scikit-Learn.
3.1.9	Подходы к построению экспериментальных исследований в области машинного обучения.
3.2	Уметь:
3.2.1	Применять методы предобработки данных (очистка, нормализация, обработка пропущенных значений, кодирование категориальных признаков).
3.2.2	Разрабатывать и реализовывать алгоритмы машинного обучения для решения задач регрессии, классификации, кластеризации.
3.2.3	Использовать Python и специализированные библиотеки (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn) для анализа, обработки и визуализации данных.
3.2.4	Выполнять оценку качества моделей машинного обучения с использованием метрик (MSE, RMSE, R ² , Accuracy, Precision, Recall, F1-score, ROC-AUC).
3.2.5	Применять методы кросс-валидации и подбора гиперпараметров для повышения точности моделей.
3.2.6	Реализовывать и интерпретировать методы регуляризации (Lasso, Ridge, Elastic Net) для предотвращения переобучения.
3.2.7	Работать с различными алгоритмами классификации (логистическая регрессия, SVM, деревья решений, случайные леса, градиентный бустинг).
3.2.8	Использовать методы кластеризации (K-Means, DBSCAN, иерархическая кластеризация) и анализа главных компонент (PCA) для снижения размерности данных.
3.2.9	Применять алгоритмы обработки естественного языка (NLP), включая наивный Байес, TF-IDF, Bag of Words.
3.2.10	Создавать и анализировать машинные модели на реальных данных, а также представлять результаты исследований в виде отчетов и визуализаций.
3.2.11	Разрабатывать прототипы интеллектуальных систем для автоматизации аналитических задач.
3.3	Владеть:
3.3.1	Работа с Python и его основными библиотеками (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn).
3.3.2	Использование Jupyter Notebook и Anaconda для выполнения вычислений и визуализации данных.
3.3.3	Обработка и анализ разнородных наборов данных (CSV, Excel, SQL, JSON, API-данные).
3.3.4	Методами машинного обучения:
3.3.5	Применение регрессионных алгоритмов (линейная, полиномиальная регрессия, регуляризация).
3.3.6	Использование методов классификации (логистическая регрессия, KNN, SVM, деревья решений, случайные леса).
3.3.7	Разработка моделей кластеризации (K-Means, DBSCAN, иерархическая кластеризация).
3.3.8	Применение алгоритмов ансамблирования (бустинг, случайные леса, градиентный бустинг).
3.3.9	Владение методами подбора гиперпараметров (кросс-валидация, GridSearch, RandomSearch).
3.3.10	Навыками обнаружения выбросов и работы с пропущенными значениями.
3.3.11	Методами предобработки данных и визуализации:
3.3.12	Владение методами нормализации и стандартизации данных.
3.3.13	Использование методов Feature Engineering (создание новых признаков, их отбор и преобразование).
3.3.14	Визуализация данных с помощью Seaborn и Matplotlib.
3.3.15	Основами обработки естественного языка (NLP):
3.3.16	Работа с текстовыми данными (Tokenization, Stop Words, Lemmatization, Stemming).
3.3.17	Извлечение признаков с помощью TF-IDF, Bag of Words.
3.3.18	Классификация текстов с использованием наивного Байесовского классификатора.
3.3.19	Навыками работы с моделями машинного обучения в продакшене:
3.3.20	Деплой моделей с использованием Flask/FastAPI, Docker.
3.3.21	Владение базовыми принципами DevOps для работы с моделями ML.
3.3.22	Оптимизация производительности моделей для работы с большими данными.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)								
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Пр. подг.	Примечание
	Раздел 1. Введение в курс							
1.1	Введение в машинное обучение. Области применения ML /Лек/	2	1		Л1.1 Л1.2Л2.1Л3.1			
1.2	Установка и настройка среды Anaconda, Jupyter Notebook, Python /Пр/	2	1		Л1.1 Л1.2Л2.1Л3.1			
1.3	Работа с массивами в NumPy: создание, индексация, операции /Пр/	2	1		Л1.1 Л1.2Л2.1Л3.1			
1.4	Введение /КрТО/	2	1		Л1.1 Л1.2Л2.1Л3.1			
1.5	Изучение концепций машинного обучения: основные понятия, типы обучения, области применения /Ср/	2	4		Л1.1 Л1.2Л2.1Л3.1			
1.6	Установка и настройка Anaconda, Jupyter Notebook, Python (дополнительное изучение инструментов) /Ср/	2	4		Л1.1 Л1.2Л2.1Л3.1			
	Раздел 2. Работа с библиотеками для анализа данных							
2.1	Основные библиотеки для анализа данных: NumPy, Pandas, Matplotlib, Seaborn /Лек/	2	1		Л1.1 Л1.2Л2.1Л3.1			
2.2	Подготовка данных для машинного обучения: обработка пропусков, нормализация, категоризация /Лек/	2	1		Л1.1 Л1.2Л2.1Л3.1			
2.3	Основы работы с Pandas: загрузка, очистка, предобработка данных /Пр/	2	1		Л1.1 Л1.2Л2.1Л3.1			
2.4	Визуализация данных с помощью Matplotlib и Seaborn /Пр/	2	1		Л1.1 Л1.2Л2.1Л3.1			
2.5	Дополнительные возможности NumPy: работа с многомерными массивами, функции линейной алгебры /Ср/	2	4		Л1.1 Л1.2Л2.1Л3.1			
2.6	Изучение методов предобработки данных в Pandas: работа с пропущенными значениями, группировка данных, фильтрация /Ср/	2	5		Л1.1 Л1.2Л2.1Л3.1			
2.7	Визуализация данных: изучение возможностей Seaborn и Matplotlib, построение интерактивных графиков /Ср/	2	8		Л1.1 Л1.2Л2.1Л3.1			
	Раздел 3. Основы машинного обучения							
3.1	Основы машинного обучения. Supervised и Unsupervised Learning /Лек/	2	1		Л1.1 Л1.2Л2.1Л3.1			

	Раздел 4. Регрессия							
4.1	Линейная и полиномиальная регрессия /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
4.2	Регуляризация в регрессионных моделях: Lasso, Ridge, Elastic Net /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
4.3	Реализация линейной регрессии с использованием Scikit-Learn /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
4.4	Полиномиальная регрессия и анализ переобучения моделей /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
4.5	Применение регуляризации (Lasso, Ridge) для борьбы с переобучением /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
4.6	Теоретические аспекты линейной регрессии: уравнение, MSE, R ² -метрики /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
4.7	Методы регуляризации (Lasso, Ridge, Elastic Net): принципы работы и практическое применение /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
	Раздел 5. Классификация							
5.1	Методы классификации: логистическая регрессия, KNN /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
5.2	Метод опорных векторов (SVM) и деревья решений /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
5.3	Случайные леса (Random Forest) и ансамблевые методы /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
5.4	Реализация логистической регрессии для бинарной классификации /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1	1		Работа в группах
5.5	Метод k-ближайших соседей (KNN) и его применение /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
5.6	Классификация с использованием метода опорных векторов (SVM) /Пр/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
5.7	Деревья решений и случайные леса в задачах классификации /Пр/	2	2		Л1.1 Л1.2Л2.1Л3. 1			
5.8	Теоретические аспекты логистической регрессии: сигмоидальная функция, методы оптимизации /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
5.9	Изучение алгоритма k-ближайших соседей (KNN): вычисление расстояний, влияние параметра k /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
5.10	Метод опорных векторов (SVM): теория, выбор ядра, интерпретация разделяющей гиперплоскости /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
	Раздел 6. Ансамблевые методы и бустинг							
6.1	Бустинг: AdaBoost, Gradient Boosting /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			

6.2	Реализация AdaBoost и Gradient Boosting в Python /Пр/	2	2		Л1.1 Л1.2Л2.1Л3. 1			
6.3	Основные концепции ансамблевого обучения: случайные леса, их преимущества и недостатки /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
6.4	Изучение алгоритмов градиентного бустинга (AdaBoost, XGBoost, LightGBM) и их различий /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
	Раздел 7. Обучение без учителя							
7.1	Кластеризация: K-Means, DBSCAN, иерархическая кластеризация /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
7.2	Метод главных компонент (PCA) и снижение размерности данных /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
7.3	Кластеризация методом K-Means и выбор количества кластеров /Пр/	2	2		Л1.1 Л1.2Л2.1Л3. 1			
7.4	Снижение размерности данных с использованием PCA /Пр/	2	2		Л1.1 Л1.2Л2.1Л3. 1			
7.5	Теория кластеризации: основные алгоритмы (K-Means, DBSCAN), выбор количества кластеров /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
7.6	Метод главных компонент (PCA): математика преобразования, интерпретация главных компонент /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			
	Раздел 8. Обработка естественного языка (NLP)							
8.1	Основы обработки естественного языка (NLP). Представление текста в числовом виде (Bag of Words, TF-IDF) /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
8.2	Наивный Байесовский классификатор и его применение в NLP /Лек/	2	1		Л1.1 Л1.2Л2.1Л3. 1			
8.3	Современные тренды в машинном обучении и перспективы развития /Лек/	2	2		Л1.1 Л1.2Л2.1Л3. 1			
8.4	Анализ текстовых данных: TF-IDF и наивный Байесовский классификатор /Пр/	2	4		Л1.1 Л1.2Л2.1Л3. 1			
8.5	Изучение моделей представления текстов: Bag of Words, TF-IDF, word embeddings (Word2Vec, GloVe) /Ср/	2	8		Л1.1 Л1.2Л2.1Л3. 1			

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

ЗНАТЬ

Раздел 1. Введение в машинное обучение

Что такое машинное обучение и какие существуют его основные типы?
В чем разница между обучением с учителем, без учителя и обучением с подкреплением?
Какие этапы включает в себя процесс построения модели машинного обучения?
Какие задачи решает машинное обучение? Приведите примеры из различных областей.
Какие библиотеки Python используются для машинного обучения?

Раздел 2. Работа с данными

Какие основные структуры данных используются в NumPy?
Какие способы визуализации данных можно использовать в Matplotlib и Seaborn?
Что такое нормализация и стандартизация данных? Когда их применять?
Какие методы используются для обработки пропущенных значений в датасетах?

Раздел 3. Регрессия

Что такое линейная регрессия и где она применяется?
Как работает метод наименьших квадратов в линейной регрессии?
Что такое полиномиальная регрессия и чем она отличается от линейной?
Какие методы регуляризации используются в машинном обучении и зачем они нужны?
Как проводится оценка качества регрессионной модели?

Раздел 4. Классификация

В чем принцип работы логистической регрессии?
Как работает метод k-ближайших соседей (KNN)?
Что такое метод опорных векторов (SVM) и когда его применять?
Как работают деревья решений?
В чем преимущества и недостатки случайных лесов (Random Forest)?

Раздел 5. Ансамблевые методы и бустинг

Что такое ансамблевые методы?
Как работают алгоритмы AdaBoost и Gradient Boosting?
Какие алгоритмы градиентного бустинга существуют и в чем их особенности?
Почему случайные леса считаются ансамблевым методом?

Раздел 6. Обучение без учителя

Что такое кластеризация и какие задачи она решает?
Чем отличается кластеризация DBSCAN от K-Means?
Как работает метод главных компонент (PCA) и зачем он нужен?
В чем различие между обучением без учителя и обучением с учителем?

Раздел 7. Обработка естественного языка (NLP)

Какие существуют методы представления текстов в машинном обучении?
Что такое TF-IDF и как он используется для обработки текстов?
Какие существуют модели для обработки естественного языка?
Чем нейросетевые модели NLP отличаются от традиционных методов обработки текста?

УМЕТЬ

(Практические навыки, которые студент должен освоить, используя полученные знания)

Раздел 2. Работа с данными

Загружать и обрабатывать данные с помощью Pandas.
Применять методы предобработки данных (нормализация, стандартизация, удаление пропущенных значений).
Строить визуализации с использованием Matplotlib и Seaborn.

Раздел 3. Регрессия

Реализовывать линейную регрессию в Python.
Использовать полиномиальную регрессию для моделирования сложных зависимостей.
Применять методы регуляризации (Lasso, Ridge) для предотвращения переобучения.
Оценивать качество регрессионных моделей с использованием MSE, RMSE, R^2 .

Раздел 4. Классификация

Реализовывать логистическую регрессию для задач классификации.
Использовать метод k-ближайших соседей (KNN) на практике.
Применять метод опорных векторов (SVM) для решения задач классификации.
Строить и анализировать деревья решений.
Работать со случайными лесами (Random Forest) и анализировать их параметры.

Раздел 5. Ансамблевые методы и бустинг

Реализовывать алгоритмы AdaBoost и Gradient Boosting.
Оптимизировать гиперпараметры моделей с помощью Grid Search и Random Search.

Раздел 6. Обучение без учителя

Выполнять кластеризацию с использованием K-Means.
Применять DBSCAN для выявления плотных областей данных.
Использовать метод главных компонент (PCA) для снижения размерности данных.

Раздел 7. Обработка естественного языка (NLP)

Применять наивный Байесовский классификатор для анализа текстов.
Вычислять TF-IDF и использовать его для обработки естественного языка.
Строить базовые NLP-модели для классификации текстов.

Раздел 8. Практическое применение и оценка моделей

Оценивать качество классификационных моделей с использованием метрик (Accuracy, Precision, Recall, F1-score).

Работать с матрицей ошибок (confusion matrix) для интерпретации результатов классификации. Выполнять кросс-валидацию для оценки стабильности моделей. Определять и исправлять основные проблемы, возникающие при обучении моделей машинного обучения. ВЛАДЕТЬ (Специализированные компетенции, углубленные навыки и способность к профессиональному применению технологий машинного обучения) Раздел 2. Работа с данными Владеть методами предобработки данных и инженерии признаков для машинного обучения. Работать с большими наборами данных, включая загрузку и обработку CSV, JSON и SQL-данных. Раздел 3. Регрессия Владеть методами подбора гиперпараметров регрессионных моделей. Раздел 4. Классификация Владеть методами построения сложных ансамблевых моделей. Глубоко разбираться в механизмах работы алгоритмов SVM, деревьев решений и случайных лесов. Раздел 5. Ансамблевые методы и бустинг Владеть методами обучения градиентного бустинга (XGBoost, LightGBM, CatBoost). Раздел 6. Обучение без учителя Владеть техниками кластеризации для анализа данных без разметки. Раздел 7. Обработка естественного языка (NLP) Владеть методами обработки естественного языка (word embeddings, LSTM, Transformer-модели). Раздел 8. Практическое применение и оценка моделей Владеть методами развертывания моделей машинного обучения в продакшене. Владеть методами интерпретации и объяснения моделей машинного обучения. Владеть методами оптимизации вычислений для работы с большими данными в ML.
5.2. Темы курсовых работ (проектов)
Курсовые работы не предусмотрены
5.3. Фонд оценочных средств
Фонд оценочных средств приведен в Приложении 1.
5.4. Перечень видов оценочных средств
Практические работы, контрольные работы, билеты. Виды шкал оценивания приведены в Приложении 2.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)			
6.1. Рекомендуемая литература			
6.1.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Шолле Ф.	Глубокое обучение на Python: учебное пособие	СПб.: Питер 2018
Л1.2	Рашка С., Мирджалили В.	Python и машинное обучение. Практика с примерами на Scikit-Learn, Keras и TensorFlow	СПб.: Питер 2022
6.1.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Вейдман Сет	Глубокое обучение: легкая разработка проектов на Python	СПб.: Питер 2021
6.1.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	Джурсин Д., Бос Х.	Python для сложных задач: наука о данных и машинное обучение	СПб.: Питер 2020
6.3. Перечень информационных и образовательных технологий			
6.3.1 Компетентностно-ориентированные образовательные технологии			
6.3.1.1	В рамках РПД по дисциплине «Введение в машинное обучение» компетентностно-ориентированные образовательные технологии реализуются через интеграцию теоретических знаний с практическими навыками, направленными на формирование профессиональных компетенций студентов в области анализа данных и искусственного интеллекта. Такой подход обеспечивает подготовку специалистов, способных решать реальные задачи программной инженерии и машинного обучения. Основные аспекты применения компетентностно-ориентированных технологий в данной РПД включают:		
6.3.1.2	Фокус на формирование профессиональных компетенций.		
6.3.1.3	Программа курса структурирована таким образом, чтобы способствовать развитию у студентов системного понимания алгоритмов машинного обучения, навыков работы с современными библиотеками Python (NumPy, Pandas, Matplotlib, Scikit-Learn) и умений применять полученные знания для решения прикладных задач. Компетенции формируются на трёх уровнях:		

6.3.1.4	Знать: теоретические основы алгоритмов, методы предобработки данных, принципы построения моделей.
6.3.1.5	Уметь: практическая реализация алгоритмов машинного обучения, проведение экспериментов, оценка качества моделей.
6.3.1.6	Владеть: углублённые навыки работы с инструментами анализа данных и разработки интеллектуальных систем.
6.3.1.7	Интеграция теории и практики.
6.3.1.8	Обучение строится вокруг применения теоретических знаний на практике: лекционные блоки дополняются практическими занятиями, лабораторными работами и проектными задачами, что позволяет студентам не только осваивать базовые концепции, но и вырабатывать навыки реальной работы с данными. Такой подход способствует закреплению знаний через непосредственное применение в решении задач машинного обучения.
6.3.1.9	Проблемно-ориентированное обучение и активное вовлечение студентов.
6.3.1.10	Методика обучения ориентирована на решение практических кейсов, где студенты работают над реальными или смоделированными задачами из области машинного обучения и анализа данных. Это включает:
6.3.1.11	Работа в группах и проведение дискуссий.
6.3.1.12	Анализ кейс-стади, где необходимо самостоятельно выявить проблему, подобрать методы решения и оценить результаты.
6.3.1.13	Разработку мини-проектов, позволяющих демонстрировать освоенные компетенции.
6.3.1.14	Использование современных информационных и образовательных технологий.
6.3.1.15	В образовательном процессе активно применяются цифровые инструменты и ресурсы:
6.3.1.16	Среды разработки (Jupyter Notebook, Anaconda) для практических занятий.
6.3.1.17	Интерактивные платформы и виртуальные лаборатории для моделирования и визуализации данных.
6.3.1.18	Онлайн-ресурсы, обеспечивающие доступ к актуальной литературе, видеоматериалам и специализированным форумам для обмена опытом.
6.3.1.19	Оценка формируемых компетенций.
6.3.1.20	Контроль успеваемости студентов осуществляется посредством многоступенчатой оценки, включающей:
6.3.1.21	Тестирование и контрольные задания, направленные на проверку теоретических знаний.
6.3.1.22	Практические работы, где оценивается способность применять методы машинного обучения для решения конкретных задач.
6.3.1.23	Проектную деятельность и презентации, демонстрирующие умение интегрировать теорию с практическими навыками.
6.3.1.24	Адаптивность образовательного процесса.
6.3.1.25	Компетентностно-ориентированные технологии позволяют гибко адаптировать содержание курса в соответствии с изменяющимися требованиями рынка труда и научно-техническим прогрессом. Это обеспечивает актуальность образовательной программы и возможность дальнейшего её развития с учетом современных трендов в области машинного обучения и искусственного интеллекта.
6.3.1.26	Таким образом, компетентностно-ориентированные образовательные технологии в рамках данной РПД обеспечивают всестороннее развитие студентов, позволяя им не только получать знания, но и применять их для решения реальных задач, что является ключевым условием успешной профессиональной деятельности в сфере машинного обучения и анализа данных.
6.3.2 Перечень информационных справочных систем и программного обеспечения	
6.3.2.1	6.3.2. Перечень информационно-справочных систем и программного обеспечения
6.3.2.2	В рамках реализации РПД по дисциплине «Введение в машинное обучение» используется следующий перечень информационно-справочных систем и программного обеспечения, обеспечивающих организацию учебного процесса, доступ к актуальным образовательным ресурсам и практическую работу с данными:
6.3.2.3	Операционные системы и базовая платформа:
6.3.2.4	Windows, Linux, macOS – операционные системы, обеспечивающие поддержку необходимых программных продуктов и сред разработки.
6.3.2.5	Язык программирования и среды разработки:
6.3.2.6	Python – основной язык программирования, используемый для реализации алгоритмов машинного обучения.

6.3.2.7	Jupyter Notebook – интерактивная среда для написания и выполнения кода, документирования экспериментов и визуализации данных.
6.3.2.8	Anaconda – дистрибутив Python, включающий широкий набор библиотек и инструментов для анализа данных и машинного обучения.
6.3.2.9	Библиотеки и фреймворки для анализа данных и машинного обучения:
6.3.2.1 0	NumPy – библиотека для работы с многомерными массивами и выполнения численных вычислений.
6.3.2.1 1	Pandas – инструмент для эффективной обработки и анализа структурированных данных.
6.3.2.1 2	Matplotlib и Seaborn – средства визуализации данных, позволяющие строить графики и диаграммы.
6.3.2.1 3	Scikit-Learn – библиотека, содержащая широкий набор алгоритмов для реализации задач регрессии, классификации, кластеризации и оценки качества моделей.
6.3.2.1 4	Дополнительные библиотеки (по необходимости): TensorFlow, Keras, PyTorch для ознакомления с концепциями глубокого обучения, а также другие специализированные инструменты.
6.3.2.1 5	Информационно-справочные системы и электронные библиотеки:
6.3.2.1 6	Электронные библиотеки и базы данных (eLIBRARY.RU, SpringerLink, IEEE Xplore, Scopus) – для поиска и изучения научной литературы, актуальных исследований и методических материалов.
6.3.2.1 7	Онлайн-ресурсы и образовательные порталы – специализированные сайты и платформы, предоставляющие доступ к курсам, видео-лекциям, учебным материалам и форумам для обмена опытом (например, Coursera, edX, Udacity).
6.3.2.1 8	Системы управления обучением (LMS) и электронные образовательные ресурсы:
6.3.2.1 9	Moodle, Canvas, Blackboard – платформы для организации дистанционного обучения, размещения учебных материалов, тестирования и обратной связи.
6.3.2.2 0	Системы корпоративного обучения – внутренние порталы вуза, обеспечивающие доступ к расписанию занятий, методическим материалам и результатам контроля.
6.3.2.2 1	Инструменты для совместной работы и контроля версий:
6.3.2.2 2	Git и GitHub/GitLab – системы контроля версий для организации командной работы над проектами, обмена кодом и документирования изменений.
6.3.2.2 3	Прочее программное обеспечение:
6.3.2.2 4	Офисные пакеты (Microsoft Office, LibreOffice) – для подготовки отчетов, презентаций и документации.
6.3.2.2 5	Средства визуализации данных – дополнительные инструменты для построения интерактивных графиков и анализа результатов (например, Tableau, Power BI).
6.3.2.2 6	Данный перечень обеспечивает всестороннюю поддержку учебного процесса, позволяя студентам получать доступ к современным информационным ресурсам, осваивать практические навыки и интегрировать теоретические знания с реальными задачами в области машинного обучения.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Для обеспечения качественного обучения по дисциплине «Введение в машинное обучение» предусмотрено следующее материально-техническое обеспечение, способствующее реализации образовательных целей и задач курса:
7.2	Учебные аудитории и компьютерные классы:
7.3	Оснащённые современными компьютерными рабочими местами, обеспечивающими высокую вычислительную мощность, необходимую для выполнения практических заданий по обработке и анализу данных.
7.4	Аудитории оборудованы проекторами, интерактивными досками и мультимедийными средствами, позволяющими проводить лекционные занятия с использованием презентаций, демонстраций и интерактивных упражнений.
7.5	Лабораторное оборудование и специализированные рабочие места:
7.6	Компьютерные лаборатории, оборудованные современными ПК и ноутбуками с предустановленным ПО (Anaconda, Python, Jupyter Notebook и другими необходимыми библиотеками и инструментами для анализа данных и машинного обучения).
7.7	Наличие средств для демонстрации работы программного обеспечения и проведения практических занятий в реальном времени.
7.8	Программное обеспечение и электронные ресурсы:

7.9	Среды разработки и инструменты анализа данных: Anaconda, Jupyter Notebook, IDE для Python, а также специализированные библиотеки (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn) для реализации алгоритмов машинного обучения.
7.10	Системы управления обучением (LMS): Moodle или аналогичные платформы, позволяющие организовать дистанционное обучение, размещение учебных материалов, проведение тестирования и обеспечение обратной связи.
7.11	Доступ к информационно-справочным системам и электронным библиотекам: электронные базы данных (eLIBRARY.RU, SpringerLink, IEEE Xplore, Scopus) и образовательные порталы для поиска актуальной литературы, методических рекомендаций и научных публикаций.
7.12	Инфраструктура для проведения практических занятий и проектной деятельности:
7.13	Организация групповой работы в современных рабочих зонах с возможностью использования средств коллективного взаимодействия (системы контроля версий Git, платформы GitHub/GitLab) для совместной разработки программных решений.
7.14	Обеспечение доступа к онлайн-курсам, вебинарам и специализированным образовательным ресурсам, что позволяет студентам расширять свои знания за рамками аудиторных занятий.
7.15	Поддержка технического обслуживания и обновления оборудования:
7.16	Регулярное обновление программного обеспечения и аппаратного обеспечения компьютерных классов для обеспечения соответствия современным требованиям образовательного процесса.
7.17	Наличие технического персонала, обеспечивающего оперативное решение вопросов, связанных с настройкой и функционированием вычислительной техники и ПО.
7.18	Данное материально-техническое обеспечение позволяет обеспечить эффективное проведение занятий, реализовать современные образовательные технологии и способствует формированию у студентов практических навыков, необходимых для успешной профессиональной деятельности в области машинного обучения и анализа данных.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Технологические карты дисциплины приведены в Приложении 3.

Методические указания для обучающихся по освоению дисциплины (модуля)

Данные методические указания предназначены для студентов, проходящих дисциплину «Введение в машинное обучение». Они содержат рекомендации и практические советы по организации учебного процесса, выполнению заданий и самостоятельной работе с материалами курса. Соблюдение данных рекомендаций позволит вам более эффективно осваивать как теоретический материал, так и практические навыки, необходимые для дальнейшей профессиональной деятельности.

1. Общие рекомендации

Активное участие в учебном процессе:

Регулярно посещайте лекционные и практические занятия, участвуйте в дискуссиях и групповых проектах. Ваше активное участие поможет лучше понять сложные концепции и методы машинного обучения.

Планирование самостоятельной работы:

Разбейте изучение материала на логические блоки, составьте план изучения тем, предусмотренных учебной программой, и следуйте ему. Рекомендуется выделять время не только на чтение литературы, но и на выполнение практических заданий.

Использование информационно-справочных систем:

Обращайтесь к рекомендованным электронным библиотекам и онлайн-ресурсам для поиска дополнительной литературы, научных статей и методических материалов, которые помогут углубить понимание изучаемых тем.

Работа с современными инструментами:

Освойте использование среды разработки Jupyter Notebook, Anaconda и основных библиотек Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn). Рекомендуется проводить эксперименты и реализовывать примеры из учебного материала самостоятельно.

2. Изучение теоретического материала

Чтение и анализ учебных текстов:

Внимательно изучайте рекомендованную литературу, методические разработки и презентационные материалы, предоставленные преподавателем. Делайте пометки, выписывайте ключевые определения и формулы.

Конспектирование лекций:

Ведение подробных конспектов поможет структурировать полученную информацию и станет основой для подготовки к контрольным мероприятиям и экзаменам.

Анализ примеров и кейсов:

Изучайте примеры из реальной практики, разбирайте кейс-стади, описанные в учебных материалах. Это позволит вам

увидеть применение теоретических знаний в решении практических задач.

3. Выполнение практических заданий

Лабораторные работы и практические занятия:

Активно выполняйте все лабораторные работы и практические задания. Они предназначены для закрепления теоретического материала через практическое применение. При возникновении трудностей консультируйтесь с преподавателем или одногруппниками.

Проектная деятельность:

При выполнении мини-проектов или курсовых работ используйте систему контроля версий (Git) для отслеживания изменений в коде. Работайте в группах, если предусмотрены совместные проекты, и обменивайтесь опытом.

Регулярная практика программирования:

Практикуйтесь в написании кода на Python, решайте задачи, связанные с анализом данных и построением моделей машинного обучения. Рекомендуется самостоятельно искать дополнительные задачи и проекты для углубления практических навыков.

4. Подготовка к контрольным мероприятиям

Контрольные вопросы и тесты:

Регулярно повторяйте материал с помощью контрольных вопросов, приведённых в оценочном фонде. Подготовка к тестированию должна включать как теоретическую, так и практическую часть.

Самостоятельное решение задач:

Практикуйтесь в решении типовых задач по регрессии, классификации, кластеризации и обработке естественного языка (NLP). Используйте интерактивные среды (например, Jupyter Notebook) для проверки своих решений.

Обсуждение сложных вопросов:

Если возникают вопросы или трудности при освоении материала, не стесняйтесь обращаться к преподавателям или обсуждать проблемы на тематических форумах и в группах обучения. Совместное решение задач способствует более глубокому пониманию материала.

5. Использование информационных технологий

Онлайн-курсы и вебинары:

Дополнительно участвуйте в онлайн-курсах, вебинарах и семинарах по машинному обучению. Это поможет расширить кругозор и получить современные знания в данной области.

Интерактивные образовательные платформы:

Используйте платформы LMS (например, Moodle) для доступа к учебным материалам, тестированию и получения обратной связи. Взаимодействие с преподавателем через данные системы способствует оперативному решению возникающих вопросов.

Системы контроля версий:

При выполнении практических заданий и проектов обязательно используйте Git и GitHub/GitLab для управления кодом. Это не только упорядочит процесс работы, но и поможет в дальнейшем при разработке коллективных проектов.

6. Организация рабочего пространства

Настройка среды разработки:

Перед началом работы убедитесь, что у вас установлены все необходимые программы и библиотеки: Python, Anaconda, Jupyter Notebook, а также дополнительные пакеты, рекомендованные в учебном плане.

Документирование работы:

Ведите подробную документацию своих экспериментов, ошибок и результатов. Это поможет анализировать полученные данные, а также подготовить отчёты и презентации по итогам практической работы.

Регулярное обновление знаний:

Следите за обновлениями в сфере машинного обучения, участвуйте в профессиональных сообществах и изучайте новейшие технологии и подходы. Это позволит вам оставаться конкурентоспособными в быстро меняющемся мире технологий.

Заключение

Методические указания направлены на то, чтобы сделать процесс обучения дисциплине «Введение в машинное обучение» максимально эффективным и ориентированным на практическое применение знаний. Следуя этим рекомендациям, вы сможете систематизировать свои знания, развить необходимые профессиональные навыки и подготовиться к успешной профессиональной деятельности в области анализа данных и искусственного интеллекта. Если у вас возникнут дополнительные вопросы или потребуются консультации, обращайтесь к преподавателю или воспользуйтесь электронными ресурсами университета.