

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.06
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Машинное обучение и глубокий анализ данных

(наименование дисциплины)

по направлению подготовки
09.03.04 Программная инженерия

направленность (профиль)
Программная инженерия с применением ИИ-технологий

Форма обучения: заочная

Год набора: 2024

Общая трудоемкость: 43Е

Распределение часов дисциплины по семестрам

Семестр	4	Итого
Форма контроля	Экзамен	
Вид занятий		
Лекции	4	4
Лабораторные		
Практические		
Руководство: курсовые работы (проекты) / РГР		
Промежуточная аттестация	0,35	0,35
Контактная работа	4,35	4,35
Самостоятельная работа	131	131
Контроль	8,65	8,65
Итого	144	144

Рабочую программу составил:
Профессор института цифровых технологий, доцент, докт.техн.наук, Мкртычев С.В.
(должность, ученое звание, степень, Фамилия И.О.)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, Фамилия И.О.)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана
направления подготовки

09.03.04 Программная инженерия

Срок действия рабочей программы дисциплины до «31» августа 2031 г.

УТВЕРЖДЕНО

На заседании института цифровых технологий

(протокол заседания № 1 от «05» сентября 2025 г.).

1. Цель освоения дисциплины

Цель освоения дисциплины – формирование у обучающихся теоретических знаний и практических навыков по основам машинного обучения, а также приобретение навыков использования технологий машинного обучения (МО) для глубинного анализа данных.

2. Место дисциплины в структуре ОПОП ВО

Дисциплины и практики, на освоении которых базируется данная дисциплина: Алгоритмы и программирование на основе Python.

Дисциплины и практики, для которых освоение данной дисциплины необходимо как предшествующее: Глубокое машинное обучение, Практикум по машинному обучению и анализу данных.

3. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
ПК-10 Способен разрабатывать и применять классические модели машинного обучения для извлечения знаний из данных	ПК-10.1. Знает классические модели машинного обучения и основные методы их разработки и применения для извлечения знаний из данных	Знать: типы задач ML (классификация, регрессия, кластеризация); классические алгоритмы (линейная регрессия, логистическая регрессия, деревья решений, SVM, k-means, k-NN). Уметь: выбирать модель, адекватную типу задачи и данным. Владеть: навыками использования библиотек (Scikit-learn).
	ПК-10.2. Умеет применять классические модели машинного обучения для извлечения знаний из данных	Знать: процесс кросс-валидации, методы борьбы с переобучением. Уметь: проводить полный цикл ML-проекта: от сбора данных до интерпретации результатов модели. Владеть: навыками построения пайплайнов обработки данных и обучения моделей.
	ПК-10.3. Владеет навыками разработки классических моделей машинного обучения и применения их для извлечения знаний из данных	Знать: внутреннее устройство основных алгоритмов ML. Уметь: модифицировать и комбинировать классические алгоритмы для решения специфических задач. Владеть: навыками реализации алгоритмов "с нуля" и создания ансамблевых моделей.

4. Структура и содержание дисциплины

Модуль (раздел)	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного средства)
Модуль 1 Основы машинного обучения и глубинного анализа данных	Лек1	Тема 1. Основы машинного обучения Тема 2. Основы глубинного анализа данных (Data mining)	4	2	-	-	
	Ср	Установка и настройка среды Jupyter Notebook. Подключение библиотек	4	16	8	-	Отчет по практической работе №1
Модуль 2 Использование технологий машинного обучения для глубинного анализа данных	Лек2	Тема 3. Методы и алгоритмы кластеризации в МО Тема 4. Методы и алгоритмы классификации в МО	4	2	-	-	
	Ср	Реализация алгоритма k-means	4	16	8	-	Отчет по практической работе №2
	Ср	Реализация алгоритма иерархической кластеризации	4	16	8		Отчет по практической работе №3
	Ср	Реализация линейного классификатора	4	16	8		Отчет по практической работе №4
	Ср	Реализация алгоритма построения деревьев решений	4	16	8		Отчет по практической работе №5

Модуль (раздел)	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного средства)
	Ср	Реализация алгоритма к ближайших соседей	4	16	8		Отчет по практической работе №6
	Ср	Реализация алгоритма бустинга	4	17	6		Отчет по практической работе №7
	Ср	Разработка программы глубинного анализа	4	18	6		Отчет по практической работе №8
	Ср		4			-	
	ПА		4	0,35	-	-	
	Контроль	Экзамен	4	8,65	40	-	Итоговый тест
Итого:				144	100		

Схема расчета итогового балла: по накопительному рейтингу
Текущий рейтинг + Результат итогового теста

5. Образовательные технологии

В рамках учебного курса предусмотрены следующие образовательные технологии:

1. технология традиционного обучения: лекции и практические работы, самостоятельная работа;
2. технология проектного обучения: реализация и защита отчетов по практическим работам.

6. Методические указания по освоению дисциплины

6.1. Рекомендации по подготовке к лекционным занятиям

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет.

В ходе лекционных следует обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. Задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Обучающийся может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и выпускных квалификационных работ.

6.2. Рекомендации по подготовке к практическим занятиям

Обучающимся следует доводить каждую практическую работу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по рассмотренному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться обучающимся на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях обучающийся не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

По результатам выполнения работы составляется отчет, который при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что выполнение каждой работы должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

6.3. Рекомендации по подготовке к экзамену

Подготовка к экзамену способствует закреплению, углублению и обобщению знаний, получаемых, в процессе обучения, а также применению их к решению практических задач. Готовясь к экзамену, обучающийся ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На экзамене обучающийся демонстрирует то, что он приобрел в процессе обучения по конкретной учебной дисциплине.

Необходимо ориентировать обучающихся на систематическую подготовку к занятиям в течение семестра, что позволит использовать время экзаменационной сессии для систематизации знаний.

7. Оценочные средства

7.1. Паспорт оценочных средств

Семестр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
4	ПК-10	<i>Отчеты к самостоятельным работам №1-8</i>

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1 Примеры тестовых заданий

Задание №1

Процесс, при котором модель обучается на размеченных данных, известен как

Выберите один из 4 вариантов ответа:

1. + Обучение с учителем
2. - Обучение без учителя
3. - Полуавтоматическое обучение
4. – Самообучение

Задание №2

Какой алгоритм относится к обучению без учителя?

Выберите один из 4 вариантов ответа:

1. - Линейная регрессия
2. + К-средних (K-means)
3. - Дерево решений
4. - Метод опорных векторов (SVM)

Задание №3

Как называется задача прогнозирования непрерывной числовой величины?

Выберите один из 4 вариантов ответа:

1. - Классификация
2. + Регрессия
3. - Кластеризация
4. - Ассоциация

Задание №4

Выберите популярные библиотеки для машинного обучения на Python.

Выберите несколько из 5 вариантов ответа:

1. + Scikit-learn
2. + TensorFlow
3. + PyTorch
4. + Keras
5. - Apache Spark

Задание №5

Переобучение (Overfitting) модели происходит, когда:

Выберите один из 4 вариантов ответа:

1. - Модель слишком проста для данных
2. + Модель слишком хорошо подстроилась под обучающие данные и плохо работает на новых
3. - Модель не обучается достаточно долго
4. - В данных отсутствуют признаки

Задание №6

Какой метод используется для борьбы с переобучением в деревьях решений?

Выберите один из 4 вариантов ответа:

1. - Увеличение глубины дерева
2. + Стрижка (Pruning)
3. - Добавление новых признаков
4. - Уменьшение обучающей выборки

Задание №7

Метод k-ближайших соседей (K-NN) является примером:

Выберите один из 4 вариантов ответа:

1. - Обучения без учителя
2. + Обучения с учителем
3. - Ансамблевого метода
4. - Нейронной сети

Задание №8

Какой показатель оценивает долю правильных ответов модели?

Выберите один из 4 вариантов ответа:

1. + Accuracy (Точность)
2. - Recall (Полнота)
3. - F1-мера
4. - AUC-ROC

Задание №9

Выберите алгоритмы, основанные на идее ансамблирования.

Выберите несколько из 5 вариантов ответа:

1. + Случайный лес (Random Forest)
2. + Градиентный бустинг (Gradient Boosting)
3. - K-средних (K-means)
4. + AdaBoost
5. - Линейная регрессия

Задание №10

Нейронная сеть, используемая для обработки последовательностей, — это:

Выберите один из 4 вариантов ответа:

1. - Сверточная нейронная сеть (CNN)
2. + Рекуррентная нейронная сеть (RNN)
3. - Полносвязная нейронная сеть
- Автокодировщик (Autoencoder)

Критерии оценки за пройденный тест:

40 баллов выставляется обучающемуся, если он ответил правильно на все вопросы рандомной выборки 30 тестовых заданий;

0-39 баллов выставляется обучающемуся в зависимости от количества верных ответов

7.2.2. Пример практической работы

Практическое задание 1. Установка и настройка среды Jupyter Notebook. Подключение библиотек МО для Python.

1. Установите на своем компьютере пакет Anaconda (<https://www.anaconda.com/products/distribution>).
2. После завершения установки откройте меню Пуск > Все программы > Anaconda 3 и выберите Anaconda Navigator. Нажмите кнопку Launch под Jupyter Notebook, после этого он должен запуститься в браузере.
3. Откройте вкладку браузера с открытым Jupyter Notebook.
4. Выполните тестовое подключение библиотек МО Python.
5. Составьте отчет по практической работе

Требования к оформлению

Отчет должен содержать подробное описание (включая иллюстрации). Отчёт по практическому занятию выполняется на страницах формата А4 в электронном виде.

При оформлении отчёта используется сквозная нумерация страниц, считая титульный лист первой страницей. Номер страницы на титульном листе не ставится. Номера страницы ставятся по центру сверху.

При оформлении отчёта соблюдать следующие требования:

- Для заголовков: полужирный шрифт, 14 пт, центрированный.
- Для основного текста: нежирный шрифт, 14 пт, выравнивание по ширине.
- Во всех случаях тип шрифта – Times New Roman, отступ абзаца 1.25 см, полуторный междустрочный интервал.
- Поля: левое – 2 см, правое, верхнее и нижнее – 1 см.

Процедура оценивания

Оценка выполненной практической работы проводится по следующим критериям:

- Наличие всей существенной информации по работе
- Точность и полнота предоставляемых сведений
- Непротиворечивость приводимой информации
- Правильность интерпретаций и выводов, которые сделаны по результатам работы
- Степень достижения обучающимся поставленной цели
- Обоснованность применяемого решения
- Грамотность (содержательная) используемых формулировок

Критерии оценки за отчеты по практическим работам:

Формы текущего контроля	Критерии и нормы оценки
Отчеты по самостоятельным заданиям	Полностью выполненное и вовремя защищенный отчет – максимальный балл. За каждое невыполненное задание снимаются баллы в соответствии с заданием на практическое занятие. Просрочка на 1 неделю -коэффициент 0,75, за две -0,5, за три -0,25, за четыре и более -0 (учитывается факт сдачи).

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации

Семестр _____4_____

№ п/п	Вопросы к экзамену
1.	Дайте определение машинного обучения. Чем оно отличается от классического программирования?
2.	Что представляет собой обучение с учителем (Supervised Learning)?
3.	Что представляет собой обучение без учителя (Unsupervised Learning)?
4.	В чем разница между обучением с учителем и без учителя? Приведите примеры задач для каждого типа
5.	Что такое переобучение (Overfitting)? Опишите его причины и методы борьбы с ним.
6.	Что такое недообучение (Underfitting)? Каковы его причины и как его можно избежать?
7.	Для чего данные делят на обучающую, валидационную и тестовую выборки? Каковы типичные пропорции разделения?
8.	Что такое «сырые» данные (Raw Data) и какие основные задачи предобработки данных вы знаете?
9.	Дайте определение глубинного анализа данных (Data Mining). Какова его связь с машинным обучением?
10.	Что такое исследовательский анализ данных (Exploratory Data Analysis, EDA) и какие методы он в себя включает?
11.	Дайте определение задачи кластеризации. К какому типу обучения она относится?
12.	Опишите алгоритм К-средних (K-Means). Каковы его основные шаги и свойства?
13.	Что такое метрика расстояния в кластеризации? Приведите примеры (Евклидово, Манхэттенское).
14.	В чем заключаются сильные и слабые стороны алгоритма K-Means?
15.	Опишите принцип работы иерархической кластеризации. Что такое дендрограмма?
16.	В чем разница между агломеративной и дивизимной иерархической кластеризацией?
17.	Что такое DBSCAN? Как он определяет кластеры и чем он принципиально отличается от K-Means?
18.	Что представляет собой задача классификации?
19.	В чем заключается принцип работы и математическую основу наивного Байесовского классификатора?
20.	Опишите принцип работы дерева решений. Как выбираются признаки для разбиения в узле?
21.	Что такое энтропия и индекс Джини? Для чего они используются в деревьях решений?
22.	Какие алгоритмы построения деревьев решений вам известны?
23.	Каковы особенности алгоритма CART?
24.	В чем заключается метод опорных векторов (SVM)?
25.	Как работает логистическая регрессия?

№ п/п	Вопросы к экзамену
26.	Опишите алгоритм k-ближайших соседей (k-NN). Каковы его основные свойства?
27.	В чем заключается основное различие между задачами классификации и
28.	Что представляют собой основные метрики для оценки качества регрессионных моделей (например, MAE, MSE, R^2)?
29.	Что означает коэффициент детерминации R^2 ?
30.	Что такое градиентный спуск и как он используется для обучения моделей?
31.	Для решения каких задач используется линейная регрессия?
32.	Какие методы для прогнозирования временных рядов вы знаете, кроме регрессионных моделей?
33.	Что такое ARIMA?
34.	Как используется экспоненциальное сглаживание?
35.	Что такое регуляризируемая линейная регрессия и чем она лучше обычной?
36.	Что такое ансамблевые методы и в чем заключается их основная идея?
37.	В чем заключается различие методов бэггинга (Bagging) и бустинга (Boosting)?
38.	Каковы особенности алгоритма случайного леса (Random Forest)?
39.	В чем заключается основное преимущество случайного леса перед одним деревом решений?
40.	В чем заключается принцип работы градиентного бустинга (Gradient Boosting)?
41.	Какие популярные реализации градиентного бустинга вам известны?
42.	Что такое XGBoost?
43.	Что такое LightGBM?
44.	Что такое CatBoost?
45.	В чем заключается ключевое различие между бустингом и бэггингом с точки зрения последовательности/параллельности обучения моделей?
46.	Что такое стекинг (Stacking) и блендинг (Blending)? Чем они отличаются от бэггинга и бустинга?
47.	Как ансамблевые методы помогают бороться с переобучением?
48.	Как математически описывается модель линейной регрессии для одного и множества признаков?
49.	Что такое функция потерь (cost function) в линейной регрессии (MSE) и какова ее цель?
50.	В чем заключается разница между простой линейной и множественной линейной регрессией?
51.	Как можно интерпретировать веса (коэффициенты) в линейной регрессии?
52.	Какие допущения лежат в основе классической линейной регрессии?
53.	Дайте определение аномалии (выброса) в данных. Чем поиск аномалий отличается от кластеризации?
54.	Какие подходы к поиску аномалий вам известны (supervised, semi-supervised и unsupervised)?
55.	Как можно использовать межквартильный размах (IQR) для обнаружения выбросов в одномерных данных?
56.	Как алгоритм k-NN может быть адаптирован для задачи поиска аномалий?
57.	Опишите принцип работы Isolation Forest. В чем его ключевая идея?
58.	Чем подход Isolation Forest отличается от кластерных методов поиска аномалий?
59.	В чем заключается сложность использования supervised-подхода для поиска аномалий?

№ п/п	Вопросы к экзамену
60.	Как можно использовать алгоритм Local Outlier Factor (LOF) для поиска локальных аномалий?

7.3.2. Критерии и нормы оценки

Семестр	Форма проведения промежуточной аттестации	Критерии и нормы оценки	
4	Экзамен (по накопительному рейтингу)	«отлично»	рейтинговый балл 85-100
		«хорошо»	рейтинговый балл 70-84
		«удовлетворительно»	рейтинговый балл 55-69
		«неудовлетворительно»	рейтинговый балл 0-54

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Запечников С. В.	Основы интеллектуального анализа данных и машинного обучения: конспект лекций	учебное пособие	2022	ЭБС "IPRbooks"
2	Ракитский А.А., Дементьева К.И.	Методы машинного обучения	учебно-методическое пособие	2023	ЭБС "IPRbooks"
3	Васильев Е.П., Орешков В.И.	Интеллектуальный анализ данных в технологиях принятия решений	учебное пособие	2023	ЭБС "IPRbooks"

8.2. Дополнительная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Алетдинова А.А., Муртазина М.Ш.	Интеллектуальный анализ больших данных	учебное пособие	2023	ЭБС "IPRbooks"
2	Меликов П.И.	Python для аналитики данных. Практический курс	учебное пособие	2023	ЭБС "IPRbooks"

8.3. Перечень профессиональных баз данных и информационных справочных систем

№ пп	Наименование	Ссылка
1	Springer Nature (Полнотекстовая коллекция журналов)	https://www.springernature.com/gp/products
2	Springer eBooks (Полнотекстовая коллекция электронных книг издательства Springer Nature)	https://link.springer.com/
3	«Кодекс»	https://kodeks.ru/
4	ELIBRARY.RU (электронная библиотека научных публикаций)	http://elibrary.ru
5	"Гарант"	https://www.garant.ru/
6	"КонсультантПлюс"	https://www.consultant.ru/
7	Техэксперт	https://cntd.ru/

8.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Реквизиты договора (дата, номер, срок действия)
1	WinPro 10 RUS Upgrd OLP NL Acdmc	Договор № 757 от 04.07.2018, срок действия – бессрочно. Контракт № 1653 от 14.12.2018, срок действия – бессрочно.
2	Office Stdandard 2013 Russian OLP NL AcademicEdition	Контракт № 690 от 19.05.2015, срок действия - бессрочно
3	Python 3.11	Лицензия Python Software Foundation License (PSFL).
4	TensorFlow	Лицензия: Apache License 2.0.
5	PyTorch	Лицензия: BSD License (Open Source).
6	Scikit-learn	Лицензия: BSD License (Open Source).

8.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории)	Перечень основного оборудования
1	Компьютерный класс. Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для проведения лабораторных работ. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная	Системные блоки (HP ProDesk), мониторы (Samsung), коммутатор (D-Link), столы ученические, столы компьютерные, стулья, доска аудиторная, экран.

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории)	Перечень основного оборудования
	аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (УЛК-402).	
2	Помещение для самостоятельной работы обучающихся (УЛК-105).	Столы, стулья, стеллажи (в т.ч. выставочные) с книгами, компьютеры, мобильные рабочие места.
3	Помещение для самостоятельной работы обучающихся (УЛК-406).	Столы компьютерные, стулья, микрокомпьютеры raspberry pi 32 bit.