

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

Б1.В.10
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Разработка и применение ИИ-моделей для анализа данных
(наименование дисциплины)

по направлению подготовки

09.03.04 Программная инженерия

направленность (профиль)

Программная инженерия с применением ИИ-технологий

Форма обучения: заочная

Год набора: 2025

Общая трудоемкость: 4 ЗЕ

Распределение часов дисциплины по семестрам

Семестр	8	Итого
Форма контроля	зачет	
Вид занятий		
Лекции	4	4
Лабораторные	-	-
Практические	-	-
Руководство: курсовые работы (проекты) / РГР	-	-
Промежуточная аттестация	0,25	0,25
Контактная работа	4,25	4,25
Самостоятельная работа	136	136
Контроль	3,75	3,75
Итого	144	144

Рабочую программу составила:

Доцент института цифровых технологий, к.э.к.наук, Раченко Т.А.

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

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

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

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

09.03.04 Программная инженерия

(код и наименование направления подготовки, специальности в соответствии с ФГОС ВПО)

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 – формирование у обучающихся компетенций в области разработки и применения моделей искусственного интеллекта для анализа данных, включая все этапы жизненного цикла таких моделей: от подготовки данных до развертывания в промышленной среде.

Задачи:

1. Изучение основных типов моделей искусственного интеллекта (классические алгоритмы машинного обучения, нейронные сети, ансамблевые методы) и их применимости для различных задач анализа данных.

2. Освоение методов предварительной обработки данных, включая очистку, трансформацию, кодирование признаков и работу с несбалансированными выборками.

3. Формирование навыков выбора архитектуры модели, обучения, настройки гиперпараметров и оценки качества на основе релевантных метрик.

4. Овладение инструментальными средствами разработки ИИ-моделей (Python, Scikit-learn, TensorFlow, PyTorch, Keras) и платформами для работы с большими данными (Spark, Dask).

5. Изучение подходов к интерпретации и объяснимости моделей, а также методов мониторинга их работы в продуктивной среде.

6. Приобретение навыков развертывания обученных моделей в виде веб-сервисов и их интеграции в приложения.

2. Место дисциплины в структуре ОПОП ВО

Данная дисциплина (учебный курс) относится к части, формируемой участниками образовательных отношений.

Дисциплины, учебные курсы, на освоении которых базируется данная дисциплина (учебный курс) – Информационные системы и технологии, Программирование на Python, Базы данных и управление данными, Математический анализ, Теория вероятностей и математическая статистика, Введение в интеллектуальный анализ данных, Машинное обучение.

Дисциплины, учебные курсы, для которых необходимы знания, умения, навыки, приобретаемые в результате изучения данной дисциплины (учебного курса) – Системы искусственного интеллекта, Инженерия машинного обучения, Практикум по машинному обучению и анализу данных, выполнение выпускной квалификационной работы.

1. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
ПК-12. Способен разрабатывать и применять модели искусственного интеллекта для обработки данных	ПК-12.1. Знает модели искусственного интеллекта и методы их разработки и применения для обработки данных	Знать: – Современные типы моделей искусственного интеллекта (регрессия, классификация, кластеризация, нейросетевые архитектуры). – Этапы жизненного цикла моделей ИИ и методы обработки данных на каждом этапе. – Метрики оценки качества моделей и подходы к интерпретации результатов. – Инструментальные средства разработки ИИ (Python, Scikit-learn, TensorFlow, PyTorch) и платформы для работы с большими данными (Spark). Уметь: – Подбирать и обосновывать выбор модели ИИ в зависимости от поставленной задачи и характера данных. – Выполнять предварительную обработку данных и формировать признаки, пригодные для обучения. – Проводить обучение, настройку гиперпараметров и оценку качества моделей. – Развертывать обученную модель в виде сервиса. Владеть: – Навыками программирования на Python с использованием библиотек для анализа данных и машинного обучения. – Методами оценки и отбора признаков, работы с пропусками и выбросами. – Технологиями кросс-валидации, оптимизации гиперпараметров. – Практическими приемами развертывания моделей (REST API, контейнеризация).

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
	ПК-12.2. Умеет применять модели искусственного интеллекта для обработки данных	Знать: – Современные подходы к трансформации данных (кодирование категориальных признаков, масштабирование, агрегация). – Техники работы с текстовыми, временными рядами и изображениями как входными данными для моделей ИИ. – Основные алгоритмы оптимизации и методы регуляризации для предотвращения переобучения. Уметь: – Применять методы предобработки данных, адаптированные под конкретную задачу. – Выбирать метрики качества и интерпретировать результаты модели. – Экспериментировать с различными архитектурами моделей и подходами к обучению. – Использовать инструменты автоматизации экспериментов и отслеживания версий моделей. Владеть: – Навыками работы в интерактивных средах разработки (Jupyter Notebook). – Методами визуализации данных и результатов моделей. – Практическими приемами тонкой настройки предобученных моделей (transfer learning).
	ПК-12.3. Владеет навыками разработки моделей искусственного интеллекта и применения их для обработки данных	Знать: – Современные платформы и фреймворки для промышленного применения моделей ИИ (MLflow, TensorFlow Serving, ONNX). – Принципы построения конвейеров обработки данных и автоматизации обучения. – Методы мониторинга моделей в продуктивной среде и обнаружения дрейфа данных. Уметь: – Разрабатывать модульные и переиспользуемые компоненты для обучения и применения моделей. – Организовывать эксперименты с фиксацией гиперпараметров и метрик. – Упаковывать модель в контейнер и развертывать как микросервис. Владеть: – Навыками использования Git и систем контроля версий для управления кодом моделей. – Методами оптимизации производительности моделей при инференсе. – Технологиями CI/CD для автоматизации развертывания моделей.

4. Структура и содержание дисциплины Разработка и применение ИИ-моделей для анализа данных

5. Модуль (раздел)	Вид учеб- ной ра- боты	Наименование тем занятий (учебной ра- боты)	Се- местр	Объём, ч	Баллы	Ин- тер- ак- тив, ч	Формы те- кущего контроля
1. Основы раз- работки ИИ-мо- делей	лекция	Тема 1. Введение в разработку и применение ИИ-моделей. Классификация задач анализа данных. Методы предобработки данных (очистка, масштабирование, кодирование признаков).	8	2	–	–	–
	самост.	Изучение лекционного материала, подготовка к практическим занятиям	8	12	–	–	–
	лекция	Тема 1.1. (лекционный обзор) Классические модели машинного обучения (регрессия, классификация, кластеризация). Нейросетевые модели.	8	2	–	–	–
	самост.	Изучение теоретического материала по темам 1.1, 2, 2.1, 3, 3.1 (по учебной литературе)	8	28	–	–	–
	самост. (практ.)	Практическая работа №1. Предобработка данных и разведочный анализ (Pandas, Matplotlib).	8	8	12	–	Отчёт по прак- тической ра- боте (защита)
	самост. (практ.)	Практическая работа №2. Построение моде- лей регрессии и классификации (Scikit-learn). Оценка качества.	8	8	12	–	Отчёт по прак- тической ра- боте (защита)

5. Модуль (раздел)	Вид учеб- ной ра- боты	Наименование тем занятий (учебной ра- боты)	Се- местр	Объём, ч	Баллы	Ин- тер- ак- тив, ч	Формы те- кущего контроля
	самост. (практ.)	Практическая работа №3. Обучение нейрон- ной сети (TensorFlow/Keras) для задачи класси- фикации.	8	8	12	—	Отчёт по прак- тической ра- боте (защита)
	самост.	Выполнение заданий по практическим работам №1–3	8	24	—	—	—
2. Применение моделей для анализа данных	самост.	Изучение лекционного материала, выполнение заданий по обучению моделей, настройке ги- перпараметров	8	24	—	—	—
	самост. (практ.)	Практическая работа №4. Обработка боль- ших данных с использованием PySpark и по- строение модели в MLlib.	8	8	12	—	Отчёт по прак- тической ра- боте (защита)
	самост.	Выполнение заданий по работе с большими данными	8	8	—	—	—
3. Развертыва- ние и эксплуата- ция моделей	самост.	Изучение лекционного материала по разверты- ванию моделей, интерпретации, логированию	8	12	—	—	—
	самост. (практ.)	Практическая работа №5. Создание веб-сер- виса для предсказаний (Flask, Docker). Интер- претация моделей (SHAP/LIME).	8	8	12	—	Отчёт по прак- тической ра- боте (защита)
	самост.	Выполнение заданий по развертыванию моде- лей, интерпретации, логированию; подготовка отчётов	8	12	—	—	—

5. Модуль (раздел)	Вид учеб- ной ра- боты	Наименование тем занятий (учебной ра- боты)	Се- местр	Объём, ч	Баллы	Ин- тер- ак- тив, ч	Формы те- кущего контроля
	ПА	Зачет	8	0,25	40	—	Оценка уровня знаний и уме- ний
Итого				144	100	—	

Схема расчета итогового балла:

Текущий рейтинг формируется из суммы баллов за защиту 5 практических работ (максимум 60 баллов, по 12 баллов за каждую). Результат итогового тестирования (или устного собеседования) оценивается в 40 баллов. Итоговая оценка определяется по накопительной системе:

«зачтено» – при наборе от 55 до 100 итоговых баллов;

«не зачтено» – при наборе менее 55 баллов.

5. Образовательные технологии

В процессе изучения дисциплины используется технология традиционного обучения (лекции, практические работы, самостоятельная работа студента)

6. Методические указания по освоению дисциплины

Для успешного освоения дисциплины необходимы посещение студентами лекционных и практических занятий, самостоятельная работа студентов с лекционным материалом и учебной литературой.

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

В ходе лекционных занятий полезно задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

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

Студентам следует

- при подготовке к практическим занятиям обязательно использовать не только лекции, учебную литературу, но и другие источники;
- в начале занятий задавать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и использовании при решении задач, предложенных для самостоятельного решения;
- на занятиях доводить каждую задачу до окончательного ответа, демонстрировать понимание проведенных расчетов (рассуждений), в случае затруднений обращаться к преподавателю.

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что решение задач проводится по рассмотренному на лекциях материалу и связано, 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 студентом на практических занятиях как в результате обсуждения и анализа лекционного материала, так и в процессе решения задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (что очень важно) для активной проработки лекционного материала.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений (рассуждений, преобразований) составить краткий план решения проблемы (задачи). Решение задач следует излагать подробно, вычисления (рассуждения, преобразования) располагать в строгом порядке. Решение при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Полезно (если это возможно) решать задачу несколькими способами и сравнивать полученные результаты. Решение задач определённого типа нужно продолжать до приобретения твердых навыков в их решении.

Самостоятельная работа студентов по предмету организуется в следующих формах:

- 1) самостоятельное изучение основного теоретического материала, ознакомление с дополнительной литературой, Интернет-ресурсами;
- 2) решение профессиональных задач из реальной предметной области.

В качестве учебно-методического обеспечения самостоятельной работы используется основная и дополнительная литература по предмету, Интернет-ресурсы, материал лекций, указания, выданные преподавателем при проведении практических работ.

Подготовка к зачету способствует закреплению, углублению и систематизации знаний, получаемых в процессе обучения. Готовясь к зачету, студент ликвидирует имеющиеся пробелы в знаниях, упорядочивает свои знания. На зачете студент демонстрирует как теоретические знания, приобретённые в процессе обучения по данной учебной дисциплине, так и навыки их практического использования при решении задач.

Необходимо ориентировать студентов на систематическую подготовку к занятиям в течение семестра, поскольку это позволит освоить основы изучаемой дисциплины, а время экзаменационной сессии можно будет использовать для систематизации уже имеющихся знаний.

7. Оценочные средства

7.1 Паспорт оценочных средств экзамену

Се- местр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
	ПК-12	<i>Тестовые задания 1-300 Вопросы к зачету Практические работы 1-5</i>

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

Отчеты по практическим работам

(наименование оценочного средства)

Типовые примеры заданий

Практическая работа №1. Предобработка данных и разведочный анализ (Pandas, Matplotlib)

Цель работы: освоить основные методы очистки, преобразования и визуализации данных, необходимые для подготовки данных к обучению моделей ИИ.

Задание:

1. Загрузить датасет из открытого источника (например, Kaggle, UCI) или предоставленный преподавателем.

2. Провести первичный анализ:

- вывести структуру датасета;
- получить статистические характеристики;
- проверить наличие пропусков и дубликатов.

3. Выполнить очистку данных:

- обработать пропуски (удалить или заполнить средним/медианой/модой);
- выявить и обработать выбросы (метод IQR или Z-score);
- удалить дубликаты.

4. Выполнить кодирование категориальных признаков (One-Hot Encoding или Label Encoding в зависимости от характера признака).

5. Выполнить масштабирование числовых признаков (StandardScaler или MinMaxScaler).

6. Построить не менее трёх информативных визуализаций (гистограммы, ящики с усами, корреляционная матрица, парные графики) и сделать выводы о структуре данных.

Форма отчета: Jupyter Notebook с кодом и комментариями. Отчёт в формате PDF, содержащий скриншоты результатов, краткие выводы по каждому этапу.

Практическая работа №2. Построение моделей регрессии и классификации (Scikit-learn)

Цель работы: научиться применять классические алгоритмы машинного обучения для решения задач регрессии и классификации, оценивать качество моделей и сравнивать их эффективность.

Задание:

1. На основе данных, подготовленных в ПР №1, разделить выборку на обучающую и тестовую (например, 80/20).

2. Построить не менее двух моделей регрессии (линейная регрессия, Ridge, Lasso) для задачи прогнозирования целевой числовой переменной (если задача регрессии).

3. Построить не менее двух моделей классификации (логистическая регрессия, дерево решений, случайный лес) для задачи классификации (если задача классификации).

4. Оценить качество моделей с использованием соответствующих метрик:

- для регрессии: MAE, MSE, RMSE, R^2 ;
- для классификации: accuracy, precision, recall, F1-score.

5. Сравнить результаты разных моделей, сделать выводы о выборе наилучшей модели.

Форма отчета: Jupyter Notebook с кодом и комментариями. Отчёт в формате PDF с описанием выбранных моделей, метрик, результатов сравнения и итоговыми выводами.

Практическая работа №3. Обучение нейронной сети (TensorFlow/Keras) для задачи классификации

Цель работы: освоить создание, обучение и оценку нейросетевых моделей, научиться настраивать архитектуру сети и параметры обучения.

Задание:

1. Подготовить данные для классификации (можно использовать тот же датасет или взять известный, например, MNIST, Fashion-MNIST, Iris).

2. Построить модель полносвязной нейронной сети с использованием Keras:

- выбрать количество слоёв и нейронов;
- выбрать функцию активации для скрытых слоёв (ReLU) и выходного слоя (softmax для многоклассовой классификации).

3. Выбрать функцию потерь (categorical_crossentropy или binary_crossentropy), оптимизатор (Adam, SGD) и метрики (accuracy).

4. Обучить модель, отобразить динамику потерь и точности на обучающей и валидационной выборках.
5. Оценить качество модели на тестовой выборке.
6. Поэкспериментировать с архитектурой (добавить dropout, изменить количество слоёв) и сравнить результаты.

Форма отчета: Jupyter Notebook с кодом и комментариями. Отчёт в формате PDF с графиками обучения, таблицами результатов, сравнением архитектур и выводами.

Практическая работа №4. Обработка больших данных с использованием PySpark и построение модели в MLlib

Цель работы: познакомиться с основами распределенных вычислений на примере PySpark и применить их для обработки больших данных и обучения моделей.

Задание:

1. Загрузить датасет достаточного объёма (или искусственно увеличить имеющийся) в формате CSV/Parquet.
2. Используя PySpark DataFrame, выполнить:
 - фильтрацию строк;
 - агрегацию и группировку;
 - преобразование типов.
3. Применить преобразования для подготовки данных к обучению:
 - создать вектор признаков с помощью VectorAssembler;
 - выполнить масштабирование (StandardScaler) при необходимости.
4. Обучить модель (логистическую регрессию или дерево решений) средствами PySpark MLlib.
5. Оценить качество модели на тестовой выборке (binaryClassificationEvaluator или multiclassClassificationEvaluator).

Форма отчета: Jupyter Notebook или скрипт на Python с использованием PySpark. Отчёт в формате PDF с описанием выполненных операций, фрагментами кода, скриншотами результатов.

Практическая работа №5. Создание веб-сервиса для предсказаний (Flask, Docker) и интерпретация модели (SHAP/LIME)

Цель работы: освоить развертывание обученной модели в виде веб-сервиса и научиться интерпретировать результаты работы модели.

Задание:

1. Сохранить обученную модель из любой предыдущей работы (с помощью joblib или pickle).
2. Разработать веб-приложение на Flask, которое:
 - принимает POST-запрос с JSON-данными;
 - загружает сохранённую модель;

- возвращает предсказание в формате JSON.

3. Создать Docker-образ приложения, запустить контейнер и проверить работоспособность через curl/Postman.

4. Для выбранной модели построить объяснения с помощью SHAP (shapley values) для нескольких примеров.

5. Выполнить интерпретацию с помощью LIME для тех же примеров.

6. Настроить логирование предсказаний модели в файл (входные данные, предсказание, временная метка).

Форма отчета: Исходный код приложения, Dockerfile. Скриншоты успешных запросов и результатов интерпретации. Краткое описание шагов.

Общие требования к оформлению отчётов по практическим работам

Отчёт выполняется в текстовом редакторе, шрифт Times New Roman, 14 pt, межстрочный интервал 1,5, поля – 2 см. Объём отчёта – не менее 5 страниц (без учёта приложений). Все графические материалы (схемы, диаграммы, скриншоты) должны быть подписаны и иметь ссылки в тексте.

Код скриптов и конфигурационные файлы могут быть вынесены в приложения. Отчёт сдаётся преподавателю в электронном виде в установленный срок.

Распределение баллов по оценочным средствам (заочная форма обучения)

Оценочное средство	Максимальный балл
Практическая работа №1	12
Практическая работа №2	12
Практическая работа №3	12
Практическая работа №4	12
Практическая работа №5	12
Текущий рейтинг (сумма)	60
Зачет (итоговое тестирование / собеседование)	40
Итоговый рейтинг	100

Критерии оценивания отчётов по практическим работам

Оценка (баллы)	Критерии
12	Работа выполнена в полном объеме с соблюдением необходимой последовательности действий; отчет оформлен аккуратно, четко и без ошибок; вывод исчерпывающий и доказательный; при защите ответил на все вопросы по теме, хорошо ориентируется в материале.
8–11	Работа выполнена в полном объеме с соблюдением необходимой последовательности действий; отчет выполнен без ошибок, вывод исчерпывающий; при защите хорошо разбирается в материале, но не уверен и неполно отвечает на вопросы.
4–7	Работа выполнена не полностью, но объем выполненной части позволяет получить правильные результаты и выводы; выполнена с несущественными замечаниями; вывод по работе не раскрывает сути работы; владение понятийным аппаратом недостаточны.
1–3	Студент выполнил работу не полностью или объем выполненной части работы не позволяет сделать правильных выводов; в ответах на вопросы есть грубые ошибки; нет знания принципиальных теоретических положений темы.
0	Работа не выполнена или выполнена не по заданию.

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

Типовые примеры тестовых заданий

Задание №1

Обозначьте основные этапы жизненного цикла модели искусственного интеллекта

- а) сбор и предобработка данных
- б) обучение модели
- в) оценка качества и настройка гиперпараметров
- г) развертывание и мониторинг
- д) написание технического задания

Ответ: а, б, в, г

Задание №2

Какие методы применяются для обработки пропусков в данных?

- а) удаление строк с пропусками
- б) заполнение средним значением
- в) заполнение медианой
- г) увеличение размерности признаков

Ответ: а, б, в

Задание №3

Для кодирования категориальных признаков без упорядоченной структуры используется

- а) One-Hot Encoding
- б) Label Encoding
- в) Ordinal Encoding
- г) Binary Encoding

Ответ: а

Задание №4

Метрика, которая измеряет долю правильно классифицированных объектов

- а) accuracy
- б) precision
- в) recall
- г) F1-score

Ответ: а

Задание №5

В задаче классификации с сильным дисбалансом классов предпочтительнее использовать

- а) accuracy
- б) F1-score
- в) MAE
- г) MSE

Ответ: б

Задание №6

Какой метод регуляризации добавляет штраф, пропорциональный сумме абсолютных значений весов?

- а) Ridge (L2)
- б) Lasso (L1)
- в) ElasticNet
- г) Dropout

Ответ: б

Задание №7

Какая функция активации чаще всего используется в скрытых слоях нейронных сетей для борьбы с проблемой затухающего градиента?

- а) Sigmoid
- б) Tanh
- в) ReLU
- г) Softmax

Ответ: в

Задание №8

Алгоритм оптимизации, который адаптирует скорость обучения для каждого параметра на основе моментов

- а) SGD
- б) Momentum
- в) Adam
- г) Adagrad

Ответ: в

Задание №9

Что такое дрейф данных (data drift)?

- а) изменение статистических свойств входных данных в процессе эксплуатации модели
- б) изменение архитектуры нейронной сети
- в) переобучение модели на обучающей выборке
- г) потеря точности из-за квантования весов

Ответ: а

Задание №10

Инструмент для контейнеризации приложений

- а) Kubernetes
- б) Docker
- в) Flask
- г) MLflow

Ответ: б

Задание №11

Для развертывания модели машинного обучения в виде веб-сервиса на Python часто используется

- а) PySpark
- б) Flask
- в) TensorFlow Serving
- г) ONNX

Ответ: б

Задание №12

Метод интерпретации моделей, основанный на теории игр и вычисляющий вклад каждого признака в предсказание

- а) LIME
- б) SHAP
- в) PDP
- г) ICE

Ответ: б

Задание №13

Какая библиотека Python предназначена для распределенной обработки больших данных?

- а) PySpark
- б) Scikit-learn
- в) Dask
- г) NumPy

Ответ: а

Задание №14

Что делает метод `VectorAssembler` в PySpark MLlib?

- а) масштабирует признаки
- б) объединяет несколько столбцов в один вектор признаков

- в) кодирует категориальные признаки
- г) выполняет нормализацию данных

Ответ: б

Задание №15

Какая метрика используется для оценки качества регрессионной модели и интерпретируется как доля объясненной дисперсии?

- а) MAE
- б) MSE
- в) RMSE
- г) R^2

Ответ: г

Задание №16

Ансамблевый метод, строящий множество деревьев решений на разных подвыборках данных и усредняющий их прогнозы

- а) Gradient Boosting
- б) Random Forest
- в) AdaBoost
- г) XGBoost

Ответ: б

Задание №17

Для отслеживания экспериментов и версионирования моделей используется

- а) Docker
- б) MLflow
- в) Git
- г) Jenkins

Ответ: б

Задание №18

Какая архитектура нейронной сети наиболее эффективна для обработки изображений?

- а) RNN
- б) CNN
- в) Transformer
- г) Autoencoder

Ответ: б

Критерии оценки тестирования

Итоговое тестирование проводится в рамках зачета. Обучающемуся предлагается 30 вопросов, сформированных случайным образом из общего пула (не менее 100 заданий). Каждый правильный ответ оценивается в **1,33 балла**. Максимальный балл за тестирование – **40**.

7.3.3. Вопросы к промежуточной аттестации

Семестр 8

Вопросы к зачету

1. Какие этапы включает жизненный цикл модели искусственного интеллекта?
2. Перечислите и охарактеризуйте основные типы задач анализа данных (регрессия, классификация, кластеризация, снижение размерности).
3. Что такое признаковое пространство и как его размерность влияет на сложность модели?
4. Какие методы обработки пропусков в данных существуют? В каких случаях применяется удаление, заполнение средним, медианой, модельное заполнение?
5. Сравните One-Hot Encoding и Label Encoding: преимущества, недостатки, области применения.
6. Для чего применяется масштабирование признаков? В чем разница между стандартизацией (StandardScaler) и нормализацией (MinMaxScaler)?
7. Как выявить и обработать выбросы в данных? Приведите примеры методов (Z-score, IQR, визуализация).
8. Что такое дисбаланс классов и какие методы борьбы с ним существуют?
9. Какие инструменты и библиотеки Python используются для предобработки данных и разведочного анализа?
10. Как оценить качество исходных данных перед построением модели? Какие критерии используются?

Модуль 2. Применение моделей для анализа данных

11. В чем суть линейной регрессии? Как интерпретируются коэффициенты модели?
12. Что такое регуляризация (L1, L2)? Как она помогает предотвратить переобучение? В чем разница между Ridge и Lasso?
13. Логистическая регрессия: принцип работы, функция потерь, интерпретация вероятностей.
14. Как работают деревья решений? Какие критерии разделения используются (Gini impurity, entropy)?
15. Ансамблевые методы: случайный лес, градиентный бустинг (XGBoost, LightGBM). Их преимущества и недостатки.
16. Что такое переобучение и недообучение? Приведите признаки и способы борьбы.
17. Какие метрики используются для оценки регрессионных моделей? Поясните MAE, MSE, RMSE, R^2 .
18. Какие метрики используются для оценки классификационных моделей? Поясните accuracy, precision, recall, F1-score, ROC-AUC.
19. Что такое кросс-валидация? Какие виды кросс-валидации вы знаете? Для чего она применяется?
20. Как осуществляется подбор гиперпараметров? Сравните GridSearchCV и RandomizedSearchCV.
21. Что такое нейронная сеть прямого распространения? Из каких компонентов она состоит?
22. Какие функции активации используются в нейронных сетях? Поясните назначение ReLU, сигмоиды, tanh.
23. Объясните принцип градиентного спуска и его модификации (SGD, Adam, RMSprop).

24. Что такое обратное распространение ошибки (backpropagation)?
25. Как регуляризация применяется в нейронных сетях? Что такое dropout, batch normalization?
26. Какие архитектуры нейронных сетей используются для обработки изображений, текста, временных рядов? (CNN, RNN, LSTM, трансформеры – обзорно).
27. Что такое transfer learning и когда его целесообразно применять?
28. Как выбрать оптимальную архитектуру нейронной сети под конкретную задачу?
29. Какие фреймворки для глубокого обучения вы знаете? Сравните TensorFlow и PyTorch.
30. Какие инструменты используются для автоматизации экспериментов и отслеживания версий моделей (MLflow, Weights & Biases)?

Модуль 3. Развертывание и эксплуатация моделей

31. Какие особенности обработки больших данных необходимо учитывать при обучении моделей?
32. Что такое распределенные вычисления? В чем преимущества Apache Spark?
33. Опишите основные структуры данных PySpark: RDD, DataFrame, Dataset.
34. Как выполнить подготовку данных для обучения с использованием PySpark MLlib (VectorAssembler, StandardScaler, StringIndexer)?
35. Какие алгоритмы машинного обучения реализованы в PySpark MLlib? Приведите примеры.
36. Какие форматы хранения данных (Parquet, ORC) предпочтительны для работы с большими данными и почему?
37. Что такое конвейер (pipeline) машинного обучения в Spark MLlib?
38. Какие способы развертывания моделей машинного обучения в промышленной среде существуют?
39. Как реализовать REST API для модели машинного обучения с использованием Flask или FastAPI?
40. Что такое контейнеризация и зачем она нужна при развертывании моделей? Опишите роль Docker.
41. Какие подходы к масштабированию сервисов предсказаний вы знаете?
42. Что такое дрейф данных (data drift) и дрейф концепции (concept drift)? Как их обнаружить?
43. Какие метрики используются для мониторинга качества моделей в продуктивной среде?
44. Какие инструменты мониторинга моделей существуют? (MLflow, Prometheus, Grafana)
45. Что такое объяснимый искусственный интеллект (XAI)? Для чего он нужен?
46. Объясните принципы работы SHAP (Shapley values). Как интерпретировать SHAP-графики?
47. Что такое LIME? В чем его отличие от SHAP? Приведите примеры применения.
48. Как организовать логирование предсказаний модели и какие данные необходимо сохранять?
49. Какие требования безопасности предъявляются к API, через которые осуществляется взаимодействие с AI-сервисами?
50. Как обеспечить версионирование моделей и воспроизводимость экспериментов?

Практические кейсы:

1. Вам предоставлен набор данных с большим количеством пропусков, категориальными признаками и выбросами. Опишите последовательность действий по подготовке данных для обучения модели.
2. По результатам обучения модели классификации на тестовой выборке получены: $\text{accuracy} = 0,95$, $\text{precision} = 0,60$, $\text{recall} = 0,99$. Что можно сказать о качестве модели? Какие возможные причины такого расхождения? Предложите меры.
3. Разработанная модель машинного обучения демонстрирует высокое качество на обучающей выборке, но на новых данных качество падает. Какие шаги необходимо предпринять для выявления и устранения проблемы?
4. Для крупного интернет-магазина требуется модель прогнозирования оттока клиентов (churn prediction). Предложите архитектуру решения, начиная со сбора данных и заканчивая развертыванием модели в production.
5. Объясните, как использовать SHAP-значения для анализа влияния признаков на предсказания модели. Приведите пример из области кредитного скоринга.
6. Компания внедрила модель рекомендаций, но через несколько месяцев её качество ухудшилось. Какие возможные причины? Предложите план мониторинга и обновления модели.
7. Опишите процесс контейнеризации модели машинного обучения с использованием Docker и её развертывания в облачной среде. Какие шаги необходимо выполнить?

3. Критерии и нормы оценки

Семестр	Форма проведения промежуточной аттестации	Критерии и нормы оценки	
		Зачтено	55 и более баллов
	Зачет (по накопительному рейтингу)	Не зачтено	менее 55 баллов

8. Учебно-методическое и информационное обеспечение дисциплины

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
	Баланов, А. Н.	Машинное обучение и искусственный интеллект : учебное пособие для вузов / А. Н. Баланов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2026. — 172 с. — ISBN 978-5-507-54962-7.	Учебное пособие		Лань : электронно-библиотечная система
	Баланов, А. Н.	Искусственный интеллект. Понимание, применение и перспективы : учебник для вузов / А. Н. Баланов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2026. — 312 с. — ISBN 978-5-507-	Учебник для вузов		Лань : электронно-библиотечная система
	Алексеев, В. Е.	Структуры данных и модели вычислений : учебное пособие / В. Е. Алексеев, В. А. Таланов. — 4-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2025. — 247 с. — ISBN 978-5-4497-0939-4.	Учебное пособие		ЭБС «IPRbooks»

Дополнительная литература

№ п/п	Авторы, со- ставители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методиче- ское пособие, практикум, др.)	Год из- дания	Количество в научной биб- лиотеке / Наименова- ние ЭБС
	Арсеньева, Н. В.	Управление организацией в цифровой среде: теория, технологии, практика : монография / Н. В. Арсеньева, Л. М. Пуяткина, Г. В. Ти- хонов. — Москва : МАИ, 2025. — 209 с. — ISBN 978-5-00261-304-	Монография		Лань : элек- тронно-биб- лиотечная си- стема
	Целых, А. Н.	Применение временных рядов для анализа больших данных : учеб- ное пособие по курсу «Математические методы анализа больших данных» / А. Н. Целых, В. С. Васильев, Э. М. Котов. — Ростов-на- Дону, Таганрог : Издательство Южного федерального университета, 2021. — 84 с. — ISBN 978-5-9275-3983-3.	Учебное пособие		ЭБС

Перечень профессиональных баз данных и информационных справочных систем

№ пп	Наименование	Ссылка
	Springer Nature (Полнотекстовая коллекция журналов)	
	Springer eBooks (Полнотекстовая коллекция электронных книг издательства Springer)	
	«Кодекс»	
	Техэксперт	
	Федеральная служба по техническому и экспортному контролю	
	Kaggle (датасеты с метками безопасности)	Kaggle датасеты: полное руководство по поиску и использованию для анализа данных - DataLopata

8.4 Перечень программного обеспечения

Наименование ПО	Назначение	Лицензия
Python 3.x (Anaconda Distribution)	Язык программирования и среда для анализа данных	BSD (свободное)
Jupyter Notebook / JupyterLab	Интерактивная среда разработки	BSD
Visual Studio Code	Редактор кода	MIT (свободное)
PyCharm Community Edition	IDE для Python	Apache 2.0 (свободное)
Git	Система контроля версий	GPLv2 (свободное)
Docker Desktop Community	Контейнеризация	Apache 2.0 (свободное)
Scikit-learn	Библиотека машинного обучения	BSD
TensorFlow / Keras	Библиотека глубокого обучения	Apache 2.0
PySpark (Apache Spark)	Распределённые вычисления	Apache 2.0
MLflow	Управление жизненным циклом моделей	Apache 2.0

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	Перечень основного оборудования
	Компьютерный класс. Помещение для самостоятельной работы. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (Г-401)	Столы ученические, стулья ученические, ПК с выходом в сеть Интернет
	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 аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (УЛК-408)	Компьютер (монитор 17", системный блок Intel (R) Celeron (R) 2,66 GHz / 1 Gb / 80 Gb), маршрутизатор 2801 Router, коммутатор Catalyst, экран/интерактивная доска Smart Board TV, проектор Acer P1303W., стол преподавательский, стол ученический, стол компьютерный, стул, доска аудиторная (маркерная).
	Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (УЛК-418)	Стол ученический двухместный (моноблок), доска аудиторная 3-х секционная (меловая), стол преподавательский, стул, проектор Acer