

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: очная

Год набора: 2026

Общая трудоемкость: 4 ЗЕ

Распределение часов дисциплины по семестрам

Семестр	5	Итого
Форма контроля	зачет	
Вид занятий		
Лекции	16	16
Лабораторные	-	-
Практические	16	16
Руководство: курсовые работы (проекты) / РГР	-	-
Промежуточная аттестация	0,25	0,25
Контактная работа	32.25	32.25
Самостоятельная работа	111.75	111.75
Контроль		
Итого	144	144

Рабочую программу составила:

доцент института цифровых технологий, к.э.н.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 2030 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. Выполнение и защита выпускной квалификационной работы	ПК-12.1. Знает модели искусственного интеллекта и методы их разработки и применения для обработки данных	Знать: модели искусственного интеллекта Уметь: применять модели искусственного интеллекта для обработки данных Владеть: навыками разработки моделей искусственного интеллекта
	ПК-12.2. Умеет применять модели искусственного интеллекта для обработки данных	Знать: методы разработки моделей искусственного интеллекта Уметь: проектировать модели искусственного интеллекта для обработки данных Владеть: навыками применения моделей искусственного для обработки данных
	ПК-12.3. Владеет навыками разработки моделей искусственного интеллекта и применения их для обработки данных	Знать: методы применения для обработки данных моделей искусственного интеллекта Уметь: разрабатывать модели искусственного интеллекта для обработки данных Владеть: навыками проектирования моделей искусственного для обработки данных

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

Модуль (раздел)	Вид учебной работы	Наименование тем занятий	Семестр	Объём, ч	Баллы	Интерактив, ч	Формы текущего контроля
1. Основы разработки ИИ-моделей	лекция	Тема 1. Введение в разработку и применение ИИ-моделей. Классификация задач анализа данных.	5	2	–	–	–
	самост.	Изучение лекционного материала, подготовка к практическим занятиям	5	12	–	–	–
	лекция	Тема 1.1. Методы предобработки данных для ИИ-моделей. Очистка, масштабирование, кодирование признаков.	5	2	–	–	–
	практ.	Практическая работа №1. Предобработка данных и разведочный анализ (Pandas, Matplotlib).	5	4	20	–	Отчёт по практической работе (защита)
	самост.	Выполнение заданий по предобработке данных, подготовка отчёта	5	16	–	–	–
2. Применение моделей для анализа данных	лекция	Тема 2. Классические модели машинного обучения: регрессия, классификация, кластеризация.	5	2	–	–	–
	практ.	Практическая работа №2. Построение моделей регрессии и классификации (Scikit-learn). Оценка качества.	5	4	25	–	Отчёт по практической работе (защита)
	лекция	Тема 2.1. Нейросетевые модели. Архитектуры, обучение, регуляризация.	5	2	–	–	–

Модуль (раздел)	Вид учебной работы	Наименование тем занятий	Се-местр	Объём, ч	Баллы	Ин-терак-тив, ч	Формы теку-щего кон-троля
	практ.	Практическая работа №3. Обучение нейронной сети (TensorFlow/Keras) для задачи классификации.	5	4	25	—	Отчёт по практической работе (защита)
	самост.	Изучение лекционного материала, выполнение заданий по обучению моделей, настройке гиперпараметров	5	30	—	—	—
3. Развертывание и эксплуатация моделей	лекция	Тема 3. Работа с большими данными: распределенные вычисления с PySpark.	5	2	—	—	—
	лекция	Тема 3.1. Развертывание моделей в промышленной среде. REST API, контейнеризация (Docker).	5	2	—	—	—
	практ.	Практическая работа №4. Обработка больших данных с использованием PySpark и построение модели в MLlib.	5	2	15	—	Отчёт по практической работе (защита)
		Практическая работа №5. Создание веб-сервиса для предсказаний (Flask, Docker). Интерпретация моделей (SHAP/LIME).	5		15	—	Отчёт по практической работе (защита)
	самост.	Выполнение заданий по развертыванию моделей, интерпретации, логированию; подготовка отчётов	5	53,75	—	—	—
	пром. ат-тест.	Промежуточная аттестация	5	0,25	—	—	—

Модуль (раздел)	Вид учебной работы	Наименование тем занятий	Семестр	Объем, ч	Баллы	Интерактив, ч	Формы текущего контроля
	практик	Зачет	5	2	100	—	Итоговый тест
Итого				144	100	—	

Схема расчета итогового балла:

Текущий рейтинг (сумма баллов за практические работы) + результат итогового теста. Полученная сумма делится на 2. Максимальный итоговый балл – 100.

5. Образовательные технологии

В рамках изучения дисциплины предусмотрено использование следующих образовательных технологий:

- технология традиционного обучения;
- интерактивные технологии: учебные дискуссии (применяются во всех модулях по итогам выполнения работ).

Технологии традиционного обучения - организация учебного процесса в вузе, основанная на лекционных и практических формах обучения: объяснительно-иллюстративное обучение. Данная технология применяется во всех модулях курса.

Технология интерактивного обучения — это организация учебного процесса, которая предполагает максимальную активность обучающихся в процессе формирования ключевых компетенций. На учебной дискуссии обучающиеся представляют результаты выполнения заданной работы. Проводится обсуждение применённых решений, их эффективности и архитектуры программного кода.

6. Методические указания по освоению дисциплины

6.1 Рекомендации по подготовке к практическим занятиям

Обучающимся следует:

- при подготовке к занятиям обязательно использовать не только учебную литературу, но и другие источники (научные статьи, документацию библиотек Python, открытые датасеты);
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание путей решения поставленных задач и освоения выданных знаний, в случае затруднений обращаться к преподавателю.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если обучающийся видит несколько путей решения задачи, то нужно сравнить их и выбрать наиболее рациональный с учётом вычислительной эффективности и интерпретируемости результата. Полезно до начала решения задачи составить краткий план решения задачи. Решение проблемных задач или примеров следует излагать подробно, отделяя вспомогательные пути решения от основных. Решения при необходимости нужно сопровождать комментариями, схемами, алгоритмами.

Следует помнить, что решение каждой учебной задачи 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.

6.2 Рекомендации по подготовке к итоговой сдаче дисциплины

Подготовка к итоговой сдаче предмета способствует закреплению, углублению и обобщению знаний, получаемых в процессе обучения, а также применению их к решению практических задач. Готовясь к ней, обучающийся ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На итоговой сдаче 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 зачётной сессии для систематизации знаний.

6.2 Рекомендации по подготовке к итоговой сдаче дисциплины

Подготовка к итоговой сдаче предмета способствует закреплению, углублению и обобщению знаний, получаемых, в процессе обучения, а также применению их к решению практических задач. Готовясь к ней, обучающийся ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На итоговой сдаче 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 экзамену

Се- местр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
	ПК-12	Вопросы к зачету
		Отчеты по практическим ра- ботам №1–5

7.2 Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1 Вопросы для собеседования по модулю

Типовые примеры заданий

Модуль 1. Основные понятия и определения безопасности информации. Требования

Модуль 1. Основы разработки ИИ-моделей

1. Какие этапы включает жизненный цикл модели искусственного интеллекта?
2. Какие типы задач анализа данных вы знаете? Приведите примеры.
3. Что такое признаковое пространство и как оно влияет на сложность модели?
4. Какие методы обработки пропусков в данных существуют? В каких случаях их применяют?
5. Как работает кодирование категориальных признаков (One-Hot Encoding, Label Encoding)? В чем их преимущества и недостатки?
6. Зачем выполнять масштабирование признаков? Какие методы масштабирования вы знаете?
7. Как оценить качество данных перед построением модели?

Модуль 2. Применение моделей для анализа данных

1. В чем отличие задач регрессии, классификации и кластеризации?

2. Приведите примеры метрик для регрессии (MAE, MSE, RMSE, R^2) и классификации (accuracy, precision, recall, F1-score). В каких ситуациях каждая метрика предпочтительна?
3. Как работает кросс-валидация? Какие её разновидности существуют?
4. В чем суть переобучения и недообучения? Какие методы борьбы с ними вы знаете?
5. Как устроена простая нейронная сеть прямого распространения? Какие функции активации применяются и для чего?
6. Что такое градиентный спуск и как он используется при обучении нейронных сетей?
7. Какие подходы к настройке гиперпараметров вы знаете? Сравните GridSearchCV и RandomizedSearchCV.
8. Как регуляризация (L1, L2, dropout) помогает предотвратить переобучение?

Модуль 3. Развертывание и эксплуатация моделей

1. Какие особенности работы с большими данными необходимо учитывать при обучении моделей?
2. В чем преимущества распределенных вычислений (Spark) перед традиционными подходами?
3. Как развернуть модель машинного обучения в виде REST API? Какие фреймворки для этого используются?
4. Что такое контейнеризация и зачем она нужна при развертывании моделей?
5. Какие инструменты используются для мониторинга работы моделей в продуктивной среде?
6. Что такое дрейф данных и дрейф концепции? Как их обнаружить?
7. Какие методы интерпретации моделей (Explainable AI) вы знаете? В чем особенности SHAP и LIME?
8. Как организовать логирование предсказаний модели и для каких целей это необходимо?

Критерии оценки ответов на вопросы для собеседования

Раскрытие ответа	Баллы
90–100% ответа раскрыто, даны исчерпывающие пояснения	20
80–89% ответа раскрыто, присутствуют незначительные неточности	18
66–79% ответа раскрыто, имеются пробелы в деталях	15
50–65% ответа раскрыто, ответ поверхностный	12
Менее 50% ответа раскрыто, присутствуют принципиальные ошибки	0–11

7.2.2 Комплект отчетов по практическим работам (примеры)

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.

Критерии оценки:

Показатель	Макс. балл
Корректность загрузки и первичного анализа данных	4
Качество обработки пропусков, выбросов и дубликатов	5
Корректность кодирования категориальных признаков	4
Корректность масштабирования числовых признаков	3
Качество визуализаций и интерпретация выводов	4

Показатель	Макс. балл
Итого	20

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.

Критерии оценки:

Показатель	Макс. балл
Корректность разделения выборки	4
Выбор и реализация моделей (не менее 4 моделей)	7
Корректность оценки качества (метрики)	6
Сравнительный анализ и выводы	8
Итого	25

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.

Критерии оценки:

Показатель	Макс. балл
Корректность подготовки данных для нейросети	5
Архитектура модели и обоснование выбора	6
Процесс обучения (эпохи, батчи, валидация)	6
Оценка качества и интерпретация результатов	8
Итого	25

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.

Критерии оценки:

Показатель	Макс. балл
Корректность загрузки и работы с PySpark DataFrame	5
Выполнение операций преобразования данных	5
Применение MLlib для обучения модели	5
Итого	15

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.

Критерии оценки:

Показатель	Макс. балл
Реализация Flask-приложения	4
Правильное сохранение и загрузка модели	3
Создание рабочего Docker-образа	4
Корректное применение SHAP/LIME и интерпретация	4
Итого	15

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.

Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1 Вопросы к промежуточной аттестации (зачету)

Модуль 1. Основы разработки ИИ-моделей

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: accuracy = 0,95, precision = 0,60, 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?

Критерии оценивания зачета при прохождении итогового тестирования по БРС

Оценка по дисциплине формируется по накопительной системе балльно-рейтинговой оценки (БРС). Зачёт выставляется по результатам текущего контроля (выполнение и защита практических работ) и итогового тестирования (или устного собеседования по билетам). Максимальный итоговый балл – 100.

Процедура оценивания по накопительному рейтингу

Се- местр	Форма проведения промежуточной атте- стации	Критерии и нормы оценки
5	Зачет (по накопительному рейтингу)	«зачтено» – рейтинговый балл 55–100
		«не зачтено» – рейтинговый балл 0–54

Процедура оценивания по билетам

Если зачёт проводится в устной или письменной форме с использованием экзаменационных билетов, каждый билет содержит два теоретических вопроса и один практический кейс. Экзаменуемый должен продемонстрировать знания теоретического материала и способность применять полученные знания для решения практических задач.

Требования к ответу

Ответ должен быть научным, логически стройным и опираться на соответствующие теоретические положения, концепции, а также на практический опыт выполнения лабораторных работ. Необходимо строить ответ в единстве теории и практики, подкрепляя теоретические положения примерами.

При ответе на теоретические вопросы следует чётко формулировать определения, классификации, перечислять методы и инструменты, объяснять принципы их работы.

При решении практического кейса требуется:

- определить суть проблемы и возможные причины;
- предложить пошаговый план решения;
- обосновать выбор конкретных методов, моделей или инструментов;
- оценить эффективность предлагаемых мер.

Порядок ответа

Обучающийся самостоятельно определяет последовательность ответа на вопросы билета. Время на подготовку – 35 минут. После подготовки экзаменуемый последовательно излагает ответы. Преподаватель может задавать уточняющие и дополнительные вопросы.

Критерии оценки ответа по билетам

Оценка	Критерии
«зачтено» (55–100 баллов)	Обучающийся полностью раскрыл содержание всех вопросов билета: даны исчерпывающие, аргументированные ответы, демонстрирующие глубокое понимание материала. Практический кейс решён верно, предложены обоснованные меры, использованы профессиональные термины. Ответ логичен, грамотен, структурирован. На дополнительные вопросы даны правильные ответы.
«не зачтено» (0–54 балла)	Обучающийся не раскрыл содержание вопросов билета, допустил принципиальные ошибки, не решил практический кейс или предложил неверные решения. Не может ответить на дополнительные вопросы.

Итоговый рейтинговый балл при зачёте по накопительной системе складывается из суммы баллов за практические работы (максимум 100) и результата итогового тестирования (если предусмотрено) с последующим делением на 2, либо определяется непосредственно по сумме баллов за практические работы в зависимости от выбранной преподавателем схемы расчёта.

Применяемая схема доводится до сведения обучающихся в начале семестра.

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