

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.


« 29 » августа 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1. В.ДВ.05.02 Методы искусственного интеллекта в задачах классификации

Направление подготовки 02.03.02 Фундаментальная информатика и
информационные технологии

Профиль Современные методы машинного обучения и компьютерного зрения

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ КЛАССИФИКАЦИИ составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии.

Программу составил(а):

Приходько Татьяна Александровна, доцент, к. т. н.

Ф.И.О., должность, ученая степень, ученое звание

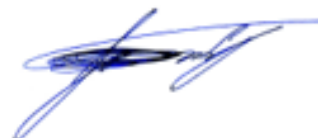


подпись

Рабочая программа дисциплины утверждена на заседании центра искусственного интеллекта

протокол № 01 «28» августа 2025 г.

Руководитель центра ИИ Коваленко А.В.



Утверждена на заседании учебно-методической комиссии факультета Компьютерных Технологий и Прикладной Математики

протокол № 1 «28» августа 2025 г.

Председатель УМК факультета Коваленко А.В.

фамилия, инициалы  подпись

Рецензенты:

Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

1. Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины «Методы искусственного интеллекта в задачах классификации» является формирование у студентов систематизированных знаний, практических умений и навыков применения современных методов искусственного интеллекта, машинного обучения для решения задач классификации данных в различных предметных областях.

Дисциплина направлена на развитие способности выбирать, реализовывать, оценивать и интерпретировать модели классификации.

1.2 Задачи дисциплины

1. Изучение теоретических основ задач классификации в машинном обучении;
2. Освоение базовых и современных алгоритмов классификации (логистическая регрессия, SVM, деревья решений, байесовский классификатор, ансамбли, нейронные сети);
3. Приобретение практических навыков предобработки данных, проектирования, обучения, оценки и оптимизации моделей классификации с использованием современных инструментов (R, Python, scikit-learn, PyTorch/TensorFlow/Keras);
4. Развитие умений анализировать результаты классификации, выбирать метрики качества, интерпретировать работу моделей;
5. Формирование навыков применения методов классификации для решения прикладных задач.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Методы искусственного интеллекта в задачах классификации» относится к части, формируемой участниками образовательных отношений Блока 1 "Дисциплины (модули) по выбору" учебного плана (Б1.В.ДВ.05.02), является альтернативой для дисциплины Б1.В.ДВ.05.01 «Анализ данных машинного обучения».

Дисциплина изучается в 7-м семестре. Для успешного освоения необходимы знания, полученные в дисциплинах: «Алгебра и введение в тензорный анализ», «Теория вероятностей и математическая статистика», «Многомерный статистический анализ и машинное обучение», «Программирование».

Преподавание ведется в виде лабораторных занятий с использованием интерактивных методов. Лабораторные работы направлены на практическое освоение методов и инструментов классификации на реальных данных.

Дисциплина формирует компетенции, необходимые для выполнения выпускной квалификационной работы и профессиональной деятельности в области вычислительных технологий.

1.4 Профессиональные роли в структуре образовательной программы

Роль 1: Data Engineer (Инженер по данным)

Задачи:

1. Проектирование и построение ETL-процессов
2. Создание и оптимизация хранилищ данных
3. Обеспечение качества и доступности данных
4. Настройка инфраструктуры для обработки больших данных
5. Интеграция разрозненных источников данных
6. Работа с данными в области природопользования, медицины, связи и телекоммуникаций

Роль 2: ML Engineer (Инженер МО)

Задачи:

1. Реализация ML-моделей в продуктивных системах
1. Оптимизация производительности и масштабирование моделей
1. Разработка ML-пайплайнов и автоматизация процессов
1. Мониторинг качества моделей в продуктиве
1. Интеграция ML-решений с бизнес-приложениями

Роль 3: MLOps (Специалист по эксплуатации ИИ)

Задачи:

- 1 Автоматизация процессов обучения и развертывания моделей
1. Мониторинг производительности ML-систем
1. Управление версиями моделей и данных
1. Обеспечение CI/CD для ML-проектов
1. Оптимизация вычислительных ресурсов

1.5 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

BD-1	Способен осуществлять поиск, сбор, очистку и предварительный анализ данных (П)
BD-1.1	Обосновывает способы и варианты применения методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи
	Знает методы заполнения пропусков в данных и удаления выбросов в табличных данных (случайные величины) Имеет навыки (умеет) очистки зашумленных временных рядов и изображений. Обнаруживает и устраняет выбросы в данных временных рядов. Владеет подходами к заполнению пропусков в данных временных рядов и изображений.
BD-1.2	Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ
	Знает основные методы понижения размерности Умеет применить основные методы понижения размерности и подбирает оптимальную размерность в зависимости от необходимой доли объяснённой дисперсии. Владеет методологией применения существующих библиотек, реализующих методы понижения размерности.
BD-1.3	Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных
	Знает и умеет применить основы методов отбора признаков и выбирает оптимальное подмножество признаков. Владеет методологией применения существующих библиотек, реализующих методы отбора признаков.
BD-1.4	Знает и умеет применить методы отбора признаков.

	Владеет способностью применять методы отбора признаков данных, значимых для исследования.
	Умеет отбирать признаки данных, значимые для исследования, Владеет методами finetuning
BD-2	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных
BD-2.1	Знает, как сформировать требования для набора данных. Владеет умениями по формированию требований к наборам и качеству данных для решения задач машинного обучения
BD-2.2	Знает приемы и инструменты для сбора данных из разрозненных источников. Умеет работать с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность. Владеет языками и инструментами для сбора данных и оценки их корректности.
LLM-2	Способен дообучать, адаптировать и оптимизировать генеративные модели под специфические задачи и условия применения
LLM-2.1	Понимает принципы fine-tune Знает: основные подходы к тонкой настройке: полная настройка всех параметров, поэтапная разморозка слоев, методы эффективной тонкой настройки (P-Tuning, LoRA, QLoRA, Adapter). Гиперпараметры, критически важные для fine-tune: learning rate, scheduler, batch size, и их отличия от обучения с нуля. Умеет: Отличать дообучение от первичного обучения, знает базовые процедуры fine-tune , анализировать задачу и выбирать наиболее подходящий метод fine-tune (полная настройка vs. эффективные методы). Владеет: Навыком осознанного выбора стратегии fine-tune под ограничения (вычислительные ресурсы, объем данных, требования к качеству). Применяет fine-tune к предобученным моделям на новых датасетах. • Методами анализа и интерпретации процесса дообучения (использование логов, графиков потерь). • Критическим мышлением для оценки целесообразности применения fine-tune в конкретном сценарии versus использования prompt engineering или RAG.
LLM-2.2	Создаёт обучающие наборы данных. Знает: Требования к данным для fine-tune: релевантность, объем, разнообразие, качество разметки. Форматы данных для популярных фреймворков (Hugging Face, TensorFlow, PyTorch) и структур задач (текст-текст, текст-изображение, инструкции и т.д.). Методы аугментации данных (data augmentation), специфичные для генеративных моделей (e.g., back-translation для текста, модификация промптов). Принципы разбиения данных на обучающую, валидационную и тестовую выборки. Умеет: Выбирать методы с учетом требований к latency и ресурсам. собирать данные из различных источников: API, веб-скрапинг, открытые датасеты, синтетическая генерация. Очищать и преобразовывать сырые данные: удаление шума, дубликатов, нормализация текста, приведение к единому формату. Размечать данные в соответствии с поставленной задачей (e.g., составлять пары "инструкция-ответ", аннотировать изображения). Применять методы аугментации данных для увеличения размера и разнообразия обучающего набора. Владеет: Навыками работы с библиотеками и инструментами для обработки данных (Pandas, NumPy, Hugging Face Datasets). Методами обеспечения репрезентативности и сбалансированности создаваемого

	<p>набора данных.</p> <p>Технологиями создания синтетических данных для задач, где реальных данных недостаточно.</p> <p>Полным циклом подготовки данных: от сбора сырых данных до формирования готового для обучения объекта (DataLoader, Dataset)</p>
MF-4	Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области ИИ.
MF-4.1	<p>Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных.</p> <p>Применяет и выбирает методы статистического машинного обучения, учитывая особенности данных и задачи, а также объясняет различия между подходами.</p> <p>Знает основные статистические методы описательного и инференционного анализа, принципы планирования экспериментов (A/B-тесты) и базовые алгоритмы машинного обучения (линейные модели, деревья).</p> <p>Умеет применять статистические методы (проверка гипотез, анализ распределений) и алгоритмы машинного обучения для исследования данных, извлечения инсайтов и проверки рабочих гипотез.</p> <p>Владеет навыком проведения полного цикла анализа данных: от предобработки и разведочного анализа (EDA) до построения, интерпретации результатов и формирования выводов.</p>
MF-4.2	<p>Способен применять статистические методы для построения предсказательных моделей, включая методы для анализа и прогнозирования временных рядов, а также моделирования нестационарных случайных процессов.</p> <p>Строит модели динамических систем для многомерных временных рядов и полей.</p> <p>Знает математические основы и предположения регрессионных, прогнозных моделей и методов анализа временных рядов (ARIMA, экспоненциальное сглаживание, подходы к работе с нестационарностью).</p> <p>Умеет строить, обучать и валидировать предсказательные модели (регрессия, классификация, прогнозирование), включая работу с временными рядами и нестационарными процессами.</p> <p>Владеет навыком выбора и настройки модели под конкретную задачу прогнозирования, диагностики её качества и интерпретации результатов прогноза.</p>
MF-4.3	<p>Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей.</p> <p>Оценивает статистические различия моделей и алгоритмов, обучаемых на данных.</p> <p>Знает и применяет модифицированные статистические критерии, A/B тестирование. Применяет оценивание на основе модифицированных доверительных интервалов, использует Байесовские тесты.</p>
MF-7	Способен применять методы дифференциальной геометрии и топологии для формализации, анализа и интерпретации структур данных и признаков пространств, включая задачи отображения, кластеризации, обучения на многообразиях и анализа устойчивости моделей.
MF-7.1	<p>Применяет методы топологического анализа для описания глобальных свойств данных и устойчивости признаков структур.</p> <p>Знает: Узнаёт и интерпретирует базовые топологические характеристики (связность, количество компонент, размерность) в примерах и визуализациях.</p> <p>Умеет: Использовать топологические дескрипторы в качестве новых признаков для модели классификации, характеризующих глобальную форму данных.</p>

	Оценивать топологическую устойчивость признакового пространства к малым возмущениям в данных. Владеет: Навыком чтения и интуитивной интерпретации персистентных диаграмм для быстрой оценки сложности структуры данных. • Методом использования TDA как инструмента для выявления неочевидных глобальных закономерностей, не улавливаемых традиционными статистическими методами.
ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.3	Решает проблемы несбалансированных данных и оценивает качество моделей Знает проблемы несбалансированных данных методы оценивания качества моделей/ Умеет применить на практике основные метрики оценки качества для задач классификации и регрессии (Б) Применяет различные типы кросс-валидации Оценивает качество моделей с учетом bias-variance trade-off (П) Владеет продвинутыми методами работы с несбалансированными данными (SMOTE weighted learning). Настраивает кастомные метрики и функции потерь. Проводит статистический анализ значимости результатов (Э)

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зачетных единицы (72 часа), их распределение по видам работ представлено в таблице

Виды работ	Всего часов	Форма обучения очная
		7 семестр (часы)
Контактная работа, в том числе:	16,2	16,2
Аудиторные занятия (всего):	16,2	16,2
занятия лекционного типа		
лабораторные занятия	16	16
практические занятия	-	-
семинарские занятия	-	-
Иная контактная работа:	0,2	0,2
Контроль самостоятельной работы (КСР)		
Промежуточная аттестация (ИКР)	0,2	0,2
Самостоятельная работа, в том числе:	55,8	55,8
Курсовая работа/проект (КР/КП) (подготовка)	-	-
Контрольная работа	-	-
Расчётно-графическая работа (РГР) (подготовка)	-	-
Выполнение индивидуальных заданий по подготовке рефератов, сообщений, презентаций	9,8	9,8

Самостоятельная проработка и материала учебников и учебных пособий, подготовка к лабораторным занятиям		38	38
Подготовка к текущему контролю		6	6
Контроль:			
Подготовка к экзамену		-	-
Общая трудоемкость	час.	72	72
	в том числе контактная работа	16,2	16,2
	зач. ед	2	2

2.2 Содержание дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы/темы дисциплины, изучаемые в 7 семестре 4 курса очной формы обучения

№	Наименование разделов (тем)	Всего	Количество часов			
			Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
Раздел 1. Введение в классификацию						
1.	Задачи классификации в ИИ и МО. Постановка задачи. Типы признаков. Этапы решения задачи классификации.	5			1	4
2.	Основы дифференциальной геометрии и топологии в интеллектуальном анализе данных. Метрики качества классификации. Матрица ошибок, Accuracy, Precision, Recall, F1-score, ROC-AUC. Кросс-валидация. Стратегии работы с дисбалансом классов. Feature Engineering	5			1	4
Раздел 2. Базовые алгоритмы классификации. Методы визуализации						
3.	Линейные модели: Логистическая регрессия. Метод опорных векторов (SVM). Ядра. Деревья решений. Анализ качества модели. Применение методов топологического анализа	9			1	8
4.	Методы визуализации и понижения размерности как инструмент геометрического исследования. Анализ качества модели. . Применение методов топологического анализа	6			1	5
Раздел 3. Ансамблевые методы и продвинутое техники						
5.	Бэггинг. Случайный лес. Стратегии работы с дисбалансом классов. Feature Engineering. Стратегии работы с дисбалансом классов. Feature Engineering.	4			1	5
6.	Бустинг (AdaBoost, Gradient Boosting, XGBoost, LightGBM, CatBoost). Стратегии работы с дисбалансом классов. Feature Engineering.	6			1	5

Раздел 4. Нейронные сети для классификации					
7.	Многослойные перцептроны. (MLP) для классификации.	7		1	6
8.	Свёрточные нейронные сети (CNN) для классификации изображений.	7		1	6
Раздел 5. Практика и применение					
9.	Кейсы применения классификации (текст, изображения, звук, табличные данные). Интерпретация моделей (SHAP, LIME).	16,8		8	12,8
<i>ИТОГО по разделам дисциплины</i>		69,8		16	55,8
Контроль самостоятельной работы (КСР)					
Промежуточная аттестация (ИКР)		0,2			
Подготовка к текущему контролю		-			
Общая трудоемкость по дисциплине		72			

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1. Занятия лекционного типа

Занятия лекционного типа не предусмотрены учебным планом.

2.3.2. Занятия семинарского типа

Занятия семинарского типа не предусмотрены учебным планом.

2.3.3. Лабораторные работы

№	Наименование раздела (темы)	Тематика лабораторных работ	Форма текущего контроля
1.	Введение в классификацию	Знакомство с библиотеками (pandas, numpy, matplotlib, scikit-learn). Загрузка и первичный анализ набора данных. Разделение на train/test. Основы дифференциальной геометрии и топологии в интеллектуальном анализе данных. Метрики качества классификации. Матрица ошибок, Accuracy, Precision, Recall, F1-score, ROC-AUC. Кросс-валидация. Стратегии работы с дисбалансом классов. Feature Engineering	Опрос по теоретическому материалу. Отчет по лабораторной работе.
2.	Базовые алгоритмы классификации. Методы визуализации Линейные модели Метод опорных векторов (SVM) Деревья решений	Реализация и оценка модели логистической регрессии на реальных данных. Анализ метрик. Применение SVM с разными ядрами. Сравнение производительности и качества. Визуализация решающих границ (для 2D). Построение и визуализация дерева решений. Анализ важности признаков.	Опрос по теоретическому материалу. Отчет по лабораторной работе.

3.	Ансамблевые методы и продвинутые техники Бэггинг и Леса, Бустинг	Реализация и оценка модели Random Forest. Сравнение с одиночным деревом. Анализ важности признаков. Применение GradientBoosting, XGBoost/LightGBM/CatBoost. Сравнение производительности и качества с Random Forest. Настройка гиперпараметров. Работа с дисбалансом: взвешивание классов, семплирование данных. Работа с признаками: создание новых признаков, Преобразование признаков: Стандартизация, нормализация, кодирование категориальных переменных. Отбор признаков (Feature Selection): Удаление маловажных или сильно коррелирующих признаков для упрощения модели и борьбы с переобучением.	Опрос по теоретическому материалу. Отчет по лабораторной работе.
4.	Нейронные сети для классификации Нейросети: MLP Нейросети: CNN	Построение, обучение и оценка MLP для классификации табличных данных с использованием Keras/TensorFlow/PyTorch. Построение, обучение и оценка CNN для классификации изображений (набор данных CIFAR-10/MNIST). Использование трансферного обучения. Работа с дисбалансом: взвешивание классов, семплирование данных. Работа с признаками: создание новых признаков, Преобразование признаков: Стандартизация, нормализация, кодирование категориальных переменных. Отбор признаков (Feature Selection): Удаление маловажных или сильно коррелирующих признаков для упрощения модели и борьбы с переобучением.	Опрос по теоретическому материалу. Отчет по лабораторной работе.
5.	Практика и интерпретация	Решение сквозной задачи классификации (на выбор). Применение SHAP/LIME для интерпретации выбранной модели.	Опрос по теоретическому материалу. Отчет по лабораторной работе.
6-7.	Практика и интерпретация	Проект с использованием ML по заданию от индустриального партнера.	Обсуждение вопросов по проекту
8.	Подведение итогов курса	Защита проектов с использованием ML по заданию от индустриальных партнеров	Защиты проектов

2.3.4. Примерная тематика курсовых работ (проектов)

Курсовая работа не предусмотрена. В качестве курсового проекта студенты защищают инфраструктуру проекта приложения с использованием ML по заданию от индустриального партнера.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Целью самостоятельной работы студента является:

- углубление знаний, полученных в результате аудиторных занятий;
- развитие навыков самостоятельной работы;
- закрепление опыта и знаний, полученных во время лабораторных занятий.

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским занятиям	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры вычислительных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 07.05.2025 г.
2	Подготовка к лабораторным занятиям	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры вычислительных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 07.05.2025 г.
3	Подготовка к решению задач и тестов	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры вычислительных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 07.05.2025 г.
4	Подготовка и выполнение проекта с использованием ML по заданию от индустриального партнера.	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры вычислительных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 07.05.2025 г.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудио-файла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

– в форме аудио-файла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии, применяемые при освоении дисциплины (модуля)

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: выполнение лабораторных работ метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

1. Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

2. Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

3. Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

4. Информационно-коммуникационные технологии (ИКТ) - расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

- Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.
- Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

5. Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

6. Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

7. Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

8. Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

9. Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

- работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;

- проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;

- анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;

- развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

При проведении лабораторных занятий участники закрепляют пройденный материал путем обсуждения вопросов, требующих особого внимания и понимания, отвечают на вопросы преподавателя и других слушателей, осуществляют решения тестов, направленных на повторение лекционного материала и нормативных документов по изучаемой тематике, выполняют решение задач, которые способствуют развитию практических навыков в области изучаемой дисциплины.

В число видов работы, выполняемой слушателями самостоятельно, входят:

- 1) поиск и изучение литературы по рассматриваемой теме;
- 2) поиск и анализ научных статей, монографий по рассматриваемой теме.

Интерактивные образовательные технологии, используемые в аудиторных занятиях: при реализации различных видов учебной работы используются следующие образовательные технологии: дискуссии, презентации, конференции. В сочетании с внеаудиторной работой они создают дополнительные условия формирования и развития требуемых компетенций обучающихся, поскольку позволяют обеспечить активное взаимодействие всех участников. Эти методы способствуют личностно-ориентированному подходу.

Все перечисленные виды и формы учебной работы и текущего контроля направлены на формирование у обучающихся профессиональных компетенций, предусмотренных при планировании результатов обучения по дисциплине и соотнесенных с планируемыми результатами освоения образовательной программы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты и устанавливается особый порядок освоения указанной дисциплины. В образовательном процессе используются социально-активные и рефлексивные методы обучения, технологии социально-культурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе.

Вышеозначенные образовательные технологии дают наиболее эффективные результаты освоения дисциплины с позиций актуализации содержания темы занятия, выработки продуктивного мышления, терминологической грамотности и компетентности обучаемого в аспекте социально направленной позиции будущего бакалавра, и мотивации к инициативному и творческому освоению учебного материала.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Методы искусственного интеллекта в задачах классификации».

Освоение дисциплины предполагает две основные формы контроля – текущая и промежуточная аттестация.

Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы и предполагает овладение материалами лекций, литературы, программы, работу студентов в ходе проведения практических занятий, а также систематическое выполнение тестовых работ, решение практических задач и иных заданий для самостоятельной работы студентов. Данный вид контроля стимулирует у студентов стремление к систематической самостоятельной работе по изучению дисциплины. Он предназначен для оценки самостоятельной работы слушателей по решению задач, выполнению практических заданий, подведения итогов тестирования. Оценивается также активность и качество результатов практической работы на занятиях, участие в дискуссиях, обсуждениях и т.п. Индивидуальные и групповые самостоятельные, аудиторные, контрольные работы по всем темам дисциплины организованы единообразным образом. Для контроля освоения содержания дисциплины используются оценочные средства. Они направлены на определение степени сформированности компетенций.

Промежуточная аттестация студентов осуществляется в рамках завершения изучения дисциплины и позволяет определить качество усвоения изученного материала, предполагает контроль и управление процессом приобретения студентами необходимых знаний, умения и навыков, определяемых по ФГОС ВО по соответствующему направлению подготовки в качестве результатов освоения учебной дисциплины.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей:

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;
- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;
- при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4.1 Оценочные средства для текущего контроля успеваемости

4.1.1. Вопросы контрольного опроса в рамках занятий лекционного и семинарского типа

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Методы искусственного интеллекта в задачах классификации».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий, кейсов и **промежуточной аттестации** в форме вопросов и заданий к **экзамену**.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Контролируемые разделы (темы) дисциплины*	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	Введение в классификацию	BD-1, BD-2, MF-7, ML-2	<i>Лабораторная работа №1</i>	<i>Вопросы к зачету</i>
2	Базовые алгоритмы классификации. Методы визуализации Линейные модели Метод опорных	BD-1, BD-2, MF-7, ML-2	<i>Лабораторная работа №2</i>	<i>Вопросы к зачету</i>

	векторов (SVM) Деревья решений			
3	Ансамблевые методы и продвинутое техники Бэггинг и Леса, Бустинг	BD-1, BD-2, LLM-2, MF-7, ML-2	<i>Лабораторная работа №3</i>	<i>Вопросы к зачету</i>
4	Нейронные сети для классификации Нейросети: MLP Нейросети: CNN	BD-1, BD-2, LLM-2, MF-7, ML-2, MF-4	<i>Лабораторная работа №4</i>	<i>Вопросы к зачету</i>
5	Практика и интерпретация	BD-1, BD-2, LLM-2, MF-7, ML-2	<i>Лабораторная работа №5-8</i>	<i>Вопросы к зачету</i>

Показатели, критерии и шкала оценки сформированных компетенций

Соответствие продвинутому уровню освоения компетенций планируемым результатам обучения и критериям их оценивания (оценка: **зачтено**):

BD-1	Способен осуществлять поиск, сбор, очистку и предварительный анализ данных (II)
BD-1.1	Обосновывает способы и варианты применения методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи
	Знает методы заполнения пропусков в данных и удаления выбросов в табличных данных (случайные величины) Имеет навыки (умеет) очистки зашумленных временных рядов и изображений. Обнаруживает и устраняет выбросы в данных временных рядов. Владеет подходами к заполнению пропусков в данных временных рядов и изображений.
BD-1.2	Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ
	Знает основные методы понижения размерности Умеет применить основные методы понижения размерности и подбирает оптимальную размерность в зависимости от необходимой доли объяснённой дисперсии. Владеет методологией применения существующих библиотек, реализующих методы понижения размерности.
BD-1.3	Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных
	Знает и умеет применить основы методов отбора признаков и выбирает оптимальное подмножество признаков. Владеет методологией применения существующих библиотек, реализующих методы отбора признаков.
BD-1.4	Знает и умеет применить методы отбора признаков. Владеет способностью применять методы отбора признаков данных, значимых для исследования.
	Умеет отбирать признаки данных, значимые для исследования, Владеет методами finetuning

BD-2	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных
BD-2.1	Знает, как сформировать требования для набора данных. Владеет умениями по формированию требований к наборам и качеству данных для решения задач машинного обучения
BD-2.2	Знает приемы и инструменты для сбора данных из разрозненных источников. Умеет работать с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность. Владеет языками и инструментами для сбора данных и оценки их корректности.
LLM-2	Способен дообучать, адаптировать и оптимизировать генеративные модели под специфические задачи и условия применения
LLM-2.1	<p>Понимает принципы fine-tune</p> <p>Знает: основные подходы к тонкой настройке: полная настройка всех параметров, поэтапная разморозка слоев, методы эффективной тонкой настройки (P-Tuning, LoRA, QLoRA, Adapter). Гиперпараметры, критически важные для fine-tune: learning rate, scheduler, batch size, и их отличия от обучения с нуля.</p> <p>Умеет: Отличать дообучение от первичного обучения, знает базовые процедуры fine-tune, анализировать задачу и выбирать наиболее подходящий метод fine-tune (полная настройка vs. эффективные методы).</p> <p>Владеет: Навыком осознанного выбора стратегии fine-tune под ограничения (вычислительные ресурсы, объем данных, требования к качеству). Применяет fine-tune к предобученным моделям на новых датасетах.</p> <ul style="list-style-type: none"> • Методами анализа и интерпретации процесса дообучения (использование логов, графиков потерь). • Критическим мышлением для оценки целесообразности применения fine-tune в конкретном сценарии versus использования prompt engineering или RAG.
LLM-2.2	<p>Создаёт обучающие наборы данных.</p> <p>Знает: Требования к данным для fine-tune: релевантность, объем, разнообразие, качество разметки. Форматы данных для популярных фреймворков (Hugging Face, TensorFlow, PyTorch) и структур задач (текст-текст, текст-изображение, инструкции и т.д.). Методы аугментации данных (data augmentation), специфичные для генеративных моделей (e.g., back-translation для текста, модификация промптов). Принципы разбиения данных на обучающую, валидационную и тестовую выборки.</p> <p>Умеет: Выбирать методы с учетом требований к latency и ресурсам. собирать данные из различных источников: API, веб-скрапинг, открытые датасеты, синтетическая генерация. Очищать и предобрабатывать сырые данные: удаление шума, дубликатов, нормализация текста, приведение к единому формату. Размечать данные в соответствии с поставленной задачей (e.g., составлять пары "инструкция-ответ", аннотировать изображения). Применять методы аугментации данных для увеличения размера и разнообразия обучающего набора.</p> <p>Владеет: Навыками работы с библиотеками и инструментами для обработки данных (Pandas, NumPy, Hugging Face Datasets).</p> <p>Методами обеспечения репрезентативности и сбалансированности создаваемого набора данных.</p> <p>Технологиями создания синтетических данных для задач, где реальных данных недостаточно.</p>

	Полным циклом подготовки данных: от сбора сырых данных до формирования готового для обучения объекта (DataLoader, Dataset)
MF-4	Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области ИИ.
MF-4.1	<p>Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных.</p> <p>Применяет и выбирает методы статистического машинного обучения, учитывая особенности данных и задачи, а также объясняет различия между подходами.</p> <p>Знает основные статистические методы описательного и инференционного анализа, принципы планирования экспериментов (A/B-тесты) и базовые алгоритмы машинного обучения (линейные модели, деревья).</p> <p>Умеет применять статистические методы (проверка гипотез, анализ распределений) и алгоритмы машинного обучения для исследования данных, извлечения инсайтов и проверки рабочих гипотез.</p> <p>Владеет навыком проведения полного цикла анализа данных: от предобработки и разведочного анализа (EDA) до построения, интерпретации результатов и формирования выводов.</p>
MF-4.2	<p>Способен применять статистические методы для построения предсказательных моделей, включая методы для анализа и прогнозирования временных рядов, а также моделирования нестационарных случайных процессов.</p> <p>Строит модели динамических систем для многомерных временных рядов и полей.</p> <p>Знает математические основы и предположения регрессионных, прогнозных моделей и методов анализа временных рядов (ARIMA, экспоненциальное сглаживание, подходы к работе с нестационарностью).</p> <p>Умеет строить, обучать и валидировать предсказательные модели (регрессия, классификация, прогнозирование), включая работу с временными рядами и нестационарными процессами.</p> <p>Владеет навыком выбора и настройки модели под конкретную задачу прогнозирования, диагностики её качества и интерпретации результатов прогноза.</p>
MF-4.3	<p>Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей.</p> <p>Оценивает статистические различия моделей и алгоритмов, обучаемых на данных.</p> <p>Знает и применяет модифицированные статистические критерии, A/B тестирование. Применяет оценивание на основе модифицированных доверительных интервалов, использует Байесовские тесты.</p>
MF-7	Способен применять методы дифференциальной геометрии и топологии для формализации, анализа и интерпретации структур данных и признаков пространств, включая задачи отображения, кластеризации, обучения на многообразиях и анализа устойчивости моделей.
MF-7.1	<p>Применяет методы топологического анализа для описания глобальных свойств данных и устойчивости признаков структур.</p> <p>Знает: Узнаёт и интерпретирует базовые топологические характеристики (связность, количество компонент, размерность) в примерах и визуализациях.</p> <p>Умеет: Использовать топологические дескрипторы в качестве новых признаков для модели классификации, характеризующих глобальную форму данных.</p> <p>Оценивать топологическую устойчивость признаков пространства к малым возмущениям в данных. Владеет: Навыком чтения и интуитивной интерпретации персистентных диаграмм для быстрой оценки сложности структуры данных.</p>

	<ul style="list-style-type: none"> • Методом использования TDA как инструмента для выявления неочевидных глобальных закономерностей, не улавливаемых традиционными статистическими методами.
ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.3	<p>Решает проблемы несбалансированных данных и оценивает качество моделей</p> <p>Знает проблемы несбалансированных данных методы оценивания качества моделей/ Умеет применить на практике основные метрики оценки качества для задач классификации и регрессии (Б)</p> <p>Применяет различные типы кросс-валидации Оценивает качество моделей с учетом bias-variance trade-off (П)</p> <p>Владеет продвинутыми методами работы с несбалансированными данными (SMOTE weighted learning). Настраивает кастомные метрики и функции потерь.</p> <p>Проводит статистический анализ значимости результатов (Э)</p>

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

4.1.2.Примеры контрольных заданий по разделам учебной дисциплины

Линейные модели (Логистическая регрессия)

Задание:

Студенту предоставлены данные о клиентах банка: возраст, ежемесячный доход и факт одобрения кредита (целевая переменная: 1 - одобрен, 0 - не одобрен).

Выполните подготовку и очистку данных. Обучите модель логистической регрессии для прогнозирования вероятности одобрения кредита. Напишите уравнение обученной модели. Проинтерпретируйте вес, стоящий при признаке "ежемесячный доход". Как изменится вероятность одобрения кредита при увеличении дохода на 10 000 рублей? Рассчитайте точность (Accuracy) и полноту (Recall) модели на тестовой выборке. Объясните, почему в данной задаче полнота может быть важнее точности.

Метод опорных векторов (SVM)

Задание:

Дан набор данных "Ирисы Фишера", но для упрощения задачи оставлены только два класса ("setosa" и "versicolor") и два признака ("длина чашелистика" и "ширина чашелистика").

Обучите классификатор SVM с линейным ядром. Визуализируйте данные и проведите решающую гиперплоскость. Отметьте опорные векторы. Повторите обучение, используя ядро RBF. Продемонстрируйте, как изменение параметра γ влияет на форму решающей границы (проделайте для $\gamma=0.1, 1, 10$). Сделайте вывод о склонности к переобучению. Для какого типа решающей границы (линейной или RBF с оптимальным γ) получается лучший результат на тестовой выборке? Ответ обоснуйте.

Деревья решений

Задание:

Используя набор данных "Титаник" (признаки: класс билета, пол, возраст, количество родственников на борту; целевая переменная: выжил/не выжил), выполните следующее:

Вручную или с помощью библиотеки постройте дерево решений глубиной не более 3. В качестве критерия разбиения используйте прирост информации (Information Gain).

Опишите словесно 2-3 ключевых правила, которые использует ваше дерево для прогнозирования выживания пассажира.

Обучите две модели: одно дерево с неограниченной глубиной и одно с параметром `max_depth=3`. Сравните их точность на обучающей и тестовой выборках. Объясните наблюдаемую разницу.

Ансамбли: Бэггинг и Случайный лес

Задание:

Работа ведется с набором данных для прогнозирования мошеннических операций с кредитными картами. Выполните подготовку и очистку данных.

Обучите одно дерево решений и оцените его точность (Accuracy) и F1-score на тестовой выборке. **Ансамблирование:** Обучите модель Случайного леса, состоящую из 100 деревьев. Сравните ее метрики (Accuracy и F1-score) с базовым деревом.

Анализ: Постройте график важности признаков, полученный из модели Случайного леса. Назовите три наиболее значимых признака для детекции мошенничества. Объясните, почему ансамблевый метод показал себя лучше.

Ансамбли: Бустинг

Задание:

Для задачи бинарной классификации (например, оттока клиентов) требуется сравнить два ансамблированных метода. Выполните подготовку и очистку данных.

Обучите две модели: Случайный лес (бэггинг) и Градиентный бустинг (XGBoost или LightGBM). Сравните модели по метрике ROC-AUC. Какая из них показала себя лучше?

Анализ: Постройте графики зависимости метрики точности на тестовой выборке от количества базовых алгоритмов (`n_estimators`) для обеих моделей. Сделайте вывод о том, какая модель быстрее сходится (достигает максимума производительности) и почему.

Нейросети: MLP

Задание:

Вам предоставлен набор данных с изображениями рукописных цифр (MNIST). Каждое изображение имеет размер 28x28 пикселей.

Подготовка данных: Выполните подготовку и очистку данных. Преобразуйте данные изображений в одномерные векторы (784 признака). Нормализуйте значения пикселей. Разбейте данные на обучающую и тестовую выборки.

Построение модели: Создайте и обучите многослойный перцептрон (MLP) со следующей архитектурой: Входной слой: 784 нейрона. Два скрытых слоя: 128 нейронов с функцией активации ReLU каждый. Выходной слой: 10 нейронов (по количеству классов) с функцией активации softmax.

Обучите модель, используя категориальную кросс-энтропию в качестве функции потерь. Постройте графики изменения функции потерь и точности на обучающей и валидационной выборках по эпохам. Сделайте вывод об отсутствии переобучения. Оцените итоговую точность на тестовой выборке.

Задачи для сбора и предобработки данных

Задача 1: "Сбор и обработка данных из API социальной сети"

Описание задачи:

Необходимо собрать данные о пользователях из API социальной сети и подготовить их для анализа поведения пользователей.

Сбор данных:

Напишите скрипт для получения данных пользователей через API

Соберите информацию о 1000 пользователей (id, имя, дата регистрации, количество друзей, количество постов). Соберите дополнительные данные о активностях пользователей за последние 30 дней.

Обработка сырых данных:

Объедините данные из разных API endpoints в единую таблицу. Обработайте вложенные JSON структуры (списки друзей, истории активностей). Сохраните собранные данные в SQLite базу данных

Предобработка:

Нормализуйте имена пользователей (приведите к единому регистру, удалите лишние пробелы). Рассчитайте производные признаки:

activity_rate (посты в день)

registration_period (дней с момента регистрации)

friends_growth_rate (скорость роста числа друзей)

Работа с пропусками:

Разработайте стратегию обработки пропущенных данных о активностях

Используйте интерполяцию для заполнения пропусков в временных рядах

Создайте флаги для обозначения импутированных значений

Задача 2 (для проектной деятельности): "Подготовка текстовых данных из внешних источников"

Описание задачи:

Подготовьте данные для модели прогнозирования цен на недвижимость, объединив данные из разных источников.

Источники данных: Открытые источники с объявлениями о продаже квартир
API карт для получения геоданных, внешние CSV файлы с инфраструктурой районов
Текстовые описания из объявлений.

Задания:

Сбор дополнительных данных:

Напишите скрипт для получения расстояний до метро через Geocoding API. Соберите данные о количестве парков, школ, магазинов в радиусе 1 км. Извлеките данные о ремонте из текстовых описаний с помощью регулярных выражений

Объединение данных:

Создайте единый датафрейм, объединив данные из 2-3 источников. Реализуйте различные типы JOIN'ов для сохранения максимального количества наблюдений. Обработайте случаи расхождения в ключах объединения

Обработка текстовых данных: Примените очистку текста (удаление стоп-слов, лемматизация).

Извлеките признаки из описаний:

Наличие ремонта (капитальный, косметический, без ремонта)

Упоминания особенностей (вид из окна, мебель, техника)

Создайте мешок слов для ключевых характеристик

Создание сложных признаков:

Рассчитайте price_per_sqm (цена за кв. метр)

Создайте признак infrastructure_score на основе собранных геоданных

Постройте признаки, описывающие соотношение комнат и площади

Создайте временные признаки (сезонность продаж)

Валидация данных:

Напишите тесты для проверки качества данных:

Проверка на дубликаты

Валидация диапазонов значений

Проверка согласованности связанных полей

Создайте пайплайн предобработки, который можно применять к новым данным

Критерии оценивания контрольных работ:

- **Полнота выполнения (все пункты задания)**
- **Качество кода (читаемость, эффективность, комментарии)**
- **Обработка крайних случаев (выбросы, пропуски, аномалии)**
- **Документирование (описание принятых решений)**
- **Воспроизводимость (возможность запуска на других машинах)**

«неудовлетворительно» – 1–2 балла – испытывает трудности применения теоретических знаний к решению практических задач; допускает принципиальные ошибки в выполнении заданий;

«удовлетворительно» – 2–3 баллов – применяет теоретические знания к решению заданий в контрольной задаче; справляется с выполнением типовых практических задач по известным алгоритмам, правилам, методам;

«хорошо» – 4 балла – правильно применяет теоретические знания к решению заданий в контрольной задаче; выполняет типовые практические задания на основе адекватных методов, способов, приемов, решает задания повышенной сложности, допускает незначительные отклонения;

«отлично» – 5 баллов – творчески применяет знания теории к решению заданий в контрольной задаче, находит оптимальные решения для выполнения практического задания; свободно выполняет типовые практические задания на основе адекватных методов, способов, приемов; решает задания повышенной сложности, находит нестандартные решения в проблемных ситуациях.

Перечень компетенций, проверяемых оценочным средством: BD-1, BD-2, ЦМ-2, MF-7, ML-2

Пример лабораторной работы

Лабораторная работа 3: Ансамбли: Бэггинг и Леса

Тема: Сравнение эффективности бэггинга и случайного леса на различных наборах данных

Цель работы: Освоить принципы работы ансамблевых методов. Научиться применять бэггинг и случайный лес на практике. Сравнить эффективность отдельных моделей и их ансамблей.

Ожидаемый результат

Реализованные модели: Дерево решений, бэггинг над деревьями, случайный лес

Сравнительный анализ: Таблица с метриками качества для всех моделей

Визуализации: Графики зависимости качества от количества базовых моделей

Анализ: Выводы о преимуществах ансамблевых методов и практических рекомендациях

Ход работы

Часть 1: Подготовка данных

Задачи:

- Загрузите набор данных "Вина" из sklearn.datasets
- Разделите данные на обучающую и тестовую выборки (70/30)
- Проведите базовую предобработку (масштабирование при необходимости)

Критерии выполнения: Созданы переменные X_train, X_test, y_train, y_test

Размеры выборок соответствуют заданному соотношению

Часть 2: Базовые модели

Задачи:

- Обучите одиночное дерево решений с максимальной глубиной 5
- Обучите бэггинг из 50 деревьев с максимальной глубиной 5
- Обучите случайный лес из 50 деревьев с максимальной глубиной 5
- Рассчитайте ассурасу для всех моделей на тестовой выборке

Критерии выполнения: Получены предсказания для всех трех моделей, рассчитаны и записаны метрики ассурасу

Часть 3: Анализ влияния размера ансамбля

Задачи:

- Исследуйте зависимость качества от количества деревьев в ансамбле (от 1 до 100)
- Постройте график ассурасу vs n_estimators для бэггинга и случайного леса
- Определите оптимальное количество деревьев для каждой модели

Критерий выполнения:

- Построен график с двумя кривыми (бэггинг и случайный лес)
- Определены точки "насыщения" качества

Часть 4: Анализ важности признаков

Задачи:

- Получите важность признаков из обученной модели случайного леса; Визуализируйте топ-5 самых важных признаков;
- Сравните с важностью признаков в одиночном дереве.

Критерии выполнения: Построен bar plot с важностью признаков, проинтерпретированы результаты.

Часть 5: Сравнительный анализ

Задачи:

- Создайте сводную таблицу с метриками всех моделей;
- Проанализируйте стабильность моделей на разных подвыборках;
- Сформулируйте выводы о применимости каждой модели.

Критерий выполнения:

Представлена сравнительная таблица, оформлен итоговый отчет.

Требуемые разделы отчета:

Теоретическая справка - краткое описание методов (1-2 абзаца на метод).

Экспериментальная часть - код и результаты выполнения всех частей работы.

Визуализации - графики из частей 3 и 4.

Анализ результатов - ответы на вопросы:

1. Какая модель показала наилучшее качество и почему?
2. Как размер ансамбля влияет на качество предсказаний?
3. Какие признаки оказались наиболее важными для классификации?
4. В каких практических задачах стоит использовать каждый из методов?

Критерии оценки:

Качество кода (20 баллов) - читаемость, комментарии, эффективность

Полнота выполнения (30 баллов) - все части работы выполнены

Визуализации (20 баллов) - качество и информативность графиков

Анализ и выводы (30 баллов) - глубина анализа, практические рекомендации

Максимальный балл: 100

*Перечень компетенций, проверяемых оценочным средством: **BD-1, BD-2, LLM-2, MF-7, ML-2***

Зачетно-экзаменационные материалы для промежуточной аттестации (зачет)

Тема 1: Введение в классификацию

Базовые понятия:

1. Дайте определение классификации как задачи машинного обучения. Чем она отличается от регрессии?
2. Что такое обучающая и тестовая выборки? Зачем нужно разделение данных?
3. Объясните понятия: признаки, метки классов, решающая граница.

Метрики качества:

1. Какие метрики качества вы знаете для задач классификации? (Accuracy, Precision, Recall, F1-score)
2. В чем разница между Precision и Recall? В каких задачах важнее каждый из этих показателей?
4. Что показывает ROC-AUC кривая? Как интерпретировать значение AUC? Как интерпретировать значение $AUC-ROC = 0.5$ и 1.0 ?
5. Почему точность (Accuracy) может быть плохой метрикой для оценки модели в условиях дисбаланса классов? Приведите пример.
6. Объясните, что такое Precision (точность) и Recall (полнота), на примере задачи обнаружения мошеннических транзакций. В чем конфликт между этими метриками?
7. Что показывает F1-мера (F1-score)? В каких ситуациях она является более информативной, чем Precision и Recall по отдельности?
8. В чем разница между ROC-AUC и Precision-Recall AUC? В какой ситуации Precision-Recall AUC является более предпочтительной метрикой?

Общие принципы:

9. Что такое переобучение (overfitting) и недообучение (underfitting)? Как их диагностировать?
10. Что такое кросс-валидация и зачем она нужна? Опишите метод k-fold cross-validation.
11. Почему простое разбиение на обучающую и тестовую выборку может быть недостаточным? Что такое переобучение (overfitting) и как разбиение данных помогает с ним бороться?
12. Что такое стратифицированное разбиение (stratified split)? В каких случаях его необходимо использовать и почему?
13. Какие методы кодирования категориальных признаков вы знаете? В чем разница между One-Hot Encoding и Label Encoding? Когда какой метод предпочтительнее?

Основы дифференциальной геометрии и топологии в интеллектуальном анализе данных

14. Объясните "гипотезу многообразия" (Manifold Hypothesis) простыми словами. Приведите пример из реальной жизни.
15. Что такое intrinsic (внутренняя) размерность данных? Почему она важна и как она связана с "проклятием размерности"?

16. Какие базовые топологические инварианты можно использовать для описания структуры данных? Что такое связность, и как она связана с кластеризацией?
17. Как методы снижения размерности, такие как PCA, t-SNE и UMAP, связаны с геометрией и топологией данных? Какой из них лучше сохраняет глобальную структуру, а какой — локальную?
18. Для чего используется алгоритм UMAP и в чем его основное топологическое преимущество перед t-SNE?

Кросс-валидация

19. Объясните принцип работы k-кратной кросс-валидации (k-fold cross-validation). Каковы ее преимущества перед простым hold-out разбиением?
20. Что такое стратифицированная k-кратная кросс-валидация? Почему она важна для несбалансированных данных?
21. В чем разница между параметрами модели и гиперпараметрами? Как кросс-валидация используется для подбора гиперпараметров?

Стратегии работы с дисбалансом классов

18. Какие существуют три основных подхода к борьбе с дисбалансом классов (на уровне данных, на уровне алгоритма и ансамблирование)?
19. В чем разница между Random Undersampling и Random Oversampling? Каковы главные недостатки каждого из этих "наивных" методов?
20. Опишите принцип работы алгоритма SMOTE для генерации синтетических примеров минорного класса. Какие у него есть преимущества перед простым дублированием?
21. Как работает метод взвешивания классов (class weighting)? Где именно в процессе обучения модели он применяется?
22. Какую метрику следует выбрать в качестве основной для оптимизации при работе с сильно несбалансированными данными и почему?

Feature Engineering

23. Что такое Feature Engineering и почему это один из самых важных этапов в построении ML-модели?
24. Какие вы знаете методы создания новых признаков? Приведите примеры для табличных, текстовых и временных данных.
25. Зачем нужно масштабировать и нормализовать признаки? Какие алгоритмы критически зависят от этой процедуры (приведите примеры), а какие — нет?
26. Объясните разницу между отбором признаков (feature selection) и выделением признаков (feature extraction). Какие методы относятся к каждому из этих подходов?
27. Что такое "проклятие размерности" и как Feature Engineering помогает с ним бороться?

Тема 2: Линейные модели

Логистическая регрессия:

9. Почему для классификации нельзя использовать обычную линейную регрессию?
10. Опишите принцип работы логистической регрессии. Какую функцию активации она использует и почему?
11. Что такое функция потерь log loss? Запишите ее формулу и объясните смысл.

Регуляризация:

12. Что такое регуляризация и зачем она нужна?

13. В чем разница между L1 (Lasso) и L2 (Ridge) регуляризацией?
14. Как регуляризация влияет на веса модели и склонность к переобучению?

Интерпретация:

15. Как интерпретировать веса в логистической регрессии?
16. Что такое odds ratio и как его вычислить по коэффициентам модели?

Метод опорных векторов (SVM)

Геометрическая интерпретация:

17. В чем состоит основная идея SVM для линейно разделимых данных?
18. Что такое зазор (margin) и почему SVM стремится его максимизировать?
19. Дайте определение опорных векторов. Почему метод так называется?

Нелинейные случаи:

20. Что делать, если данные нельзя разделить линейно? Опишите ядерный трюк.
21. Какие ядра вы знаете? (линейное, полиномиальное, RBF) В каких случаях их применять?
22. Как параметр C влияет на компромисс между шириной зазора и количеством ошибок?

Практические аспекты:

23. Нужно ли масштабировать данные для SVM? Почему?
24. В каких задачах SVM показывает хорошие результаты, а в каких - плохие?

Деревья решений

Алгоритм построения:

25. Опишите жадный алгоритм построения дерева сверху вниз.
26. Что такое критерии информативности? Дайте определение энтропии и индекса Джини.
27. Как вычисляется прирост информации (Information Gain)? По какой формуле?

Гиперпараметры:

28. Какие гиперпараметры позволяют бороться с переобучением в деревьях?
29. Как глубина дерева влияет на bias и variance?
30. Что такое классификация по большинству в листьях дерева?

Преимущества и недостатки:

31. Назовите преимущества и недостатки деревьев решений.
32. В каких областях деревья решений особенно эффективны?

Тема 3: Ансамбли: Бэггинг и Леса

Основные понятия:

33. Объясните, почему объединение нескольких "слабых" моделей может дать одну "сильную".
34. В чем разница между бэггингом и бустингом?

Бэггинг:

35. Опишите два ключевых этапа бэггинга: bootstrap sampling и агрегация.
36. Почему бэггинг эффективно уменьшает дисперсию ошибки?

Случайный лес:

37. Чем случайный лес отличается от простого бэггинга над деревьями?
38. Что такое "случайное подпространство признаков" и зачем оно нужно?
39. Как вычисляется важность признаков в случайном лесу?

Практика:

40. Почему деревья в случайном лесу обычно не обрезают?
41. Как количество деревьев влияет на качество и время работы модели?

Ансамбли: Бустинг

Основная идея:

42. В чем ключевое отличие бустинга от бэггинга?
43. Какую ошибку исправляет каждая следующая модель в ансамбле?

Алгоритмы:

44. Опишите основной принцип работы AdaBoost. Что такое веса объектов?
45. Объясните идею Gradient Boosting. Какую роль играет градиентный спуск?
46. Что такое learning rate в градиентном бустинге и как он влияет на обучение?

Современные реализации:

47. Какие преимущества у XGBoost по сравнению с классическим Gradient Boosting?
48. В чем особенности LightGBM и CatBoost?

Применение:

49. В каких задачах бустинг обычно превосходит случайный лес?
50. Какие основные гиперпараметры нужно настраивать в бустинге?

Тема 4: Нейросети: MLP

Архитектура:

51. Из каких основных компонентов состоит искусственный нейрон?
52. Что такое функция активации и зачем она нужна? Назовите популярные функции.
53. Опишите архитектуру многослойного перцептрона.

Обучение:

54. Что такое прямое распространение (Forward Pass) и обратное распространение ошибки (Backpropagation)?
55. Как обновляются веса сети в процессе обучения?
56. Что такое функция потерь (loss function) в задачах классификации?

Практические аспекты:

57. Почему функция ReLU стала популярной в скрытых слоях?
58. Зачем нужна нормализация признаков перед подачей на вход нейросети?
59. Что такое эпоха (epoch) и размер батча (batch size)?

Регуляризация:

60. Какие методы регуляризации используются в нейросетях? (Dropout, BatchNorm, L2)
61. Как избежать переобучения в глубоких нейросетях?

Общие и сравнительные вопросы

62. Сравните деревья решений и линейные модели: в каких случаях выбирать каждую?
63. Когда стоит предпочесть SVM, а когда - случайный лес?
64. Как выбрать между бустингом и случайным лесом на практике?
65. В каких задачах нейросети превосходят традиционные методы, а в каких - уступают?
66. Опишите процесс полного пайплайна машинного обучения от данных до готовой модели.
67. Как вы будете выбирать модель для задачи бинарной классификации с несбалансированными классами?
68. Что такое интерпретируемость моделей и почему она важна в некоторых областях?

Методические рекомендации к сдаче зачета и критерии оценки ответа

Промежуточная аттестация традиционно служат основным средством обеспечения в учебном процессе «обратной связи» между преподавателем и обучающимся, необходимой для стимулирования работы обучающихся и совершенствования методики преподавания учебных дисциплин. Итоговой формой контроля сформированности компетенций, обучающихся по дисциплине «Методы искусственного интеллекта в задачах классификации» является зачет. Студенты обязаны сдать зачет в соответствии с расписанием и учебным планом. Зачет по дисциплине преследует цель оценить работу студента за курс, получение теоретических знаний, их прочность, развитие творческого мышления, приобретение навыков самостоятельной работы, умение применять полученные знания для решения практических

задач и является формой контроля усвоения студентом учебной программы по дисциплине, выполнения практических, контрольных, реферативных работ. Форма проведения зачета: устно. Результат сдачи зачета по прослушанному курсу должен оцениваться как итог деятельности студента в семестре, а именно – по посещаемости лекций, результатам работы на лекционных и практических занятиях, прохождения тестовых заданий, решения расчетно-графических заданий и задач, выполнения контролируемой самостоятельной работы. Студенты, прошедшие все виды испытаний, предусмотренных оценочными средствами положительно (т.е. по каждому виду оценочных средств были получены оценки «удовлетворительно», и(или) «хорошо», и(или) «отлично») выставляется «зачтено». При этом допускается на очной форме обучения пропуск не более 20% занятий, с обязательной отработкой пропущенных семинаров. Студенты, у которых количество пропусков, превышает установленную норму, не выполнившие все виды работ и неудовлетворительно работавшие в течение семестра, проходят собеседование с преподавателем, в виде устного ответа на один теоретический вопрос и решения одного расчетно-графического задания. Преподавателю предоставляется право задавать студентам дополнительные вопросы по всей учебной программе дисциплины. Результат сдачи зачета заносится преподавателем в ведомость и зачетную книжку.

Критерии оценки зачета. Оценка «зачтено» выставляется студенту, если дан полный развернутый ответ на теоретический вопрос, логически правильно изложены ответы на дополнительные вопросы; студент показал умение свободно выполнять расчетно-графическое задание, предусмотренное дисциплиной, самостоятельность решения задания и приводимых суждений; все расчеты сделаны правильно; выводы вытекают из содержания задания, предложения обоснованы, в изложении ответов нет существенных недостатков. В то же время в ответе могут присутствовать незначительные фактические ошибки в изложении материала. Оценка «не зачтено» выставляется при несоответствии ответа заданному вопросу, наличии грубых ошибок, использовании при ответе ненадлежащих источников; студент показал пробелы в знаниях основного учебного материала, значительные пробелы в знаниях теоретических компонентов программы; неумение ориентироваться в основных научных теориях и концепциях, связанных с осваиваемой дисциплиной, неточное их описание; слабое владение научной терминологией и профессиональным инструментарием; допустил принципиальные ошибки в выполнении предусмотренной дисциплиной практического задания, изложение ответа на вопросы с существенными лингвистическими и логическими ошибками.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
 - в форме электронного документа.
- Для лиц с нарушениями опорно-двигательного аппарата:
- в печатной форме,
 - в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4.3. Методические указания по организации вычислительной инфраструктуры

Требования к аппаратному и программному обеспечению рабочих мест

Аппаратные требования.

Для выполнения лабораторных работ по изучению методов ИИ в задачах классификации студентам и преподавателю необходим стационарный компьютер или ноутбук с современной конфигурацией. Рекомендуется многопроцессорный CPU, например Intel Core i3/i5/i7 не ниже 4-го поколения или аналогичный AMD Ryzen, с поддержкой многопоточности и оперативной памятью не менее 8 ГБ. Для работы с GPU-вычислениями требуется видеокарта NVIDIA для CUDA или совместимая с OpenCL/ROCm. Компьютер должен иметь стабильное подключение к сети Интернет со скоростью не ниже 5–10 Мбит/с для скачивания SDK, библиотек и обновлений программного обеспечения.

Программные требования.

На рабочих станциях должна быть установлена современная операционная система, включая Windows 10 или 11, актуальные версии macOS или дистрибутивы GNU/Linux, при этом системы должны регулярно обновляться для поддержания безопасности и совместимости с инструментами курса. Для разработки необходимы библиотеки, указанные в разделе «Инструменты и библиотеки».

Студенты также должны иметь доступ к системе контроля версий Git, интерпретатору Python версии 3.10 и выше с менеджером пакетов pip или conda для анализа результатов и построения графиков, при необходимости с установкой Jupyter Notebook/Lab. Для локального тестирования и отладки программ может использоваться Docker, при этом на Windows требуется Docker Desktop с WSL2, а на Linux и macOS платформа поддерживается напрямую. Все программное обеспечение должно быть настроено так, чтобы студенты имели доступ ко всем инструментам во время лабораторных работ, а преподаватель мог управлять инфраструктурой и контролировать результаты, включая репозитории, CI/CD и тестирование. Необходимо обеспечить разрешение исходящих подключений по HTTPS, открытые порты 80 и 443, а также наличие прав на установку программного обеспечения или взаимодействие с системным администратором для их установки.

Инструменты и библиотеки:

Категория	Python	R
Основные ML	scikit-learn	caret, tidymodels
Обработка данных	pandas, numpy	dplyr, data.table
Визуализация	matplotlib, seaborn, plotly	ggplot2, plotly
Ансамбли	xgboost, lightgbm, catboost	randomForest, xgboost, gbm
Бустинг	xgboost, lightgbm, catboost	xgboost, gbm
Деревья решений	scikit-learn	rpart
Нейросети	tensorflow, keras, pytorch	nnet, keras
SVM	scikit-learn	e1071

Категория	Python	R
Линейные модели	scikit-learn, statsmodels	glm, lm
Интерпретация моделей	shap, eli5, lime	DALEX, iml
Оптимизация гиперпараметров	optuna, scikit-optimize	mlrMBO, tune
Несбалансированные данные	imbalanced-learn	ROSE, smotefamily
Работа с текстом	nlTK, spaCy	tm, tidytext
Верификация моделей	scikit-learn	ROCR, mlbench
Пайплайны	scikit-learn	recipes
Даты и время	pandas	lubridate
Строки	pandas	stringr
Эксперименты	mlflow	MLflow

Исходные данные: готовые датасеты, данные, собранные в ходе выполнения работ.

5. Перечень основной и дополнительной учебной литературы, информационных ресурсов и технологий необходимых для освоения дисциплины

5.1 Основная литература, в том числе публикации конференций А*

1. Sun, X., Li, J., Kovalenko, A.V., Feng, W., Ou, Y. Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation //IEEE Transactions on Automation Science and Engineering, 2023, 20(3), 1735–1744, DOI: 10.1109/TASE.2022.3185071, Q1
2. Petukhova, A.V.; Kovalenko, A.V.; Ovsyannikova, A.V. Algorithm for Optimization of Inverse Problem Modeling in Fuzzy Cognitive Maps. Mathematics 2022, 10, 3452. DOI: 10.3390/math10193452, Q1
4. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." Oncotarget 8.7 (2016): 10883.
5. Kadurin, Artur, et al. "druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico." Molecular pharmaceutics 14.9 (2017): 3098-3104.
6. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." Frontiers in pharmacology 11 (2020): 565644.
7. Khrabrov, Kuzma, et al. " ∇^2 DFT: A Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials." Advances in Neural Information Processing Systems 37 (2024): 36869-36889.
8. Polykovskiy, Daniil, et al. "Entangled conditional adversarial autoencoder for de novo drug discovery." Molecular pharmaceutics 15.10 (2018): 4398-4405.
9. Haoran Xu, Philipp Koehn, and Kenton Murray. 2022. The Importance of Being Parameters: An Intra-Distillation Method for Serious Gains. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 170–183, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
10. Wai Ching Leung, Shira Wein, and Nathan Schneider. 2022. Semantic Similarity as a Window into Vector- and Graph-Based Metrics. In Proceedings of the Second Workshop on Natural Language Generation, Evaluation, and Metrics (GEM), pages 106–115, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
11. Anna Lorincz, David Graus, Dor Lavi, and Joao Lebre Magalhaes Pereira. 2022. Transfer learning for multilingual vacancy text generation. In Proceedings of the Second Workshop on

Natural Language Generation, Evaluation, and Metrics (GEM), pages 207–222, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics

12. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов : учебное пособие для студентов вузов / Ярушкина, Надежда Глебовна, Т. В. Афанасьева, И. Г. Перфильева ; Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. - М. : ФОРУМ : ИНФРА-М, 2012. - 159 с. : ил. - (Высшее образование). - Библиогр. в конце глав. - ISBN 9785819904961. - ISBN 9785160051970.

5.2. Дополнительная литература:

1. Подоплелова, Е. С. Современные методы инженерии знаний в задачах машинного обучения : учебное пособие : [16+] / Е. С. Подоплелова ; Южный федеральный университет, Инженерно-технологическая академия. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2025. – 130 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=724468> (дата обращения: 30.10.2025). – Библиогр. в кн. – ISBN 978-5-9275-4882-8. – Текст : электронный.
2. Протодьяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие : [16+] / А. В. Протодьяконов, П. А. Пылов, В. Е. Садовников. – Москва ; Вологда : Инфра-Инженерия, 2022. – 392 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=725623> (дата обращения: 30.10.2025). – Библиогр.: с. 380-383. – ISBN 978-5-9729-1006-9. – Текст : электронный.
3. Пылов, П. А. Основы работы с моделями машинного и глубокого обучения : учебное пособие : [16+] / П. А. Пылов, Р. В. Майтак, А. В. Дягилева. – Москва ; Вологда : Инфра-Инженерия, 2023. – 256 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=725673> (дата обращения: 30.10.2025). – Библиогр.: с. 247-250. – ISBN 978-5-9729-1547-7. – Текст : электронный.
4. Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие : [16+] / Т. М. Татарникова. – Москва ; Вологда : Инфра-Инженерия, 2024. – 172 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=725643> (дата обращения: 30.10.2025). – Библиогр. в кн. – ISBN 978-5-9729-1772-3. – Текст : электронный.
5. Галыгин, А.Н. Алгоритмы автоматического формирования базы правил для систем управления на нечёткой логике: дис. ... канд. тех. наук / А.Н. Галыгин. – Красноярск, 2004. – 120 с.
6. Заде, Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений / Л. Заде. - М.: Мир, 1976. - 165с.
7. Захаров, Р.Е. Разработка логико-лингвистических моделей управления и принятия решений на базе нечеткой логики: дис. ... канд. тех. наук / Р.Е. Захаров. - Владикавказ, 2004. - 168с.

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных, информационные справочные системы и конференции

Конференции А*:

1. <https://openreview.net/forum?id=FMMF1a9ifL>
2. <https://openreview.net/forum?id=ElUrNM9U8c#discussion>
3. <https://openreview.net/forum?id=JoO6mtCLHD>
4. <https://aclanthology.org/2024.findings-emnlp.760/>
5. <https://aclanthology.org/2020.coling-main.588/>
6. https://link.springer.com/chapter/10.1007/978-3-030-72113-8_30
7. https://link.springer.com/chapter/10.1007/978-3-031-42448-9_10
8. <https://aclanthology.org/2024.findings-naacl.288/>

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
7. Президентская библиотека им. Б.Н. Ельцина <https://www.prlib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. **Консультант Плюс** - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
4. Федеральный портал "Российское образование" <http://www.edu.ru/>;
5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/>;
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;

11. Образовательный портал "Учеба" <http://www.ucheba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы [http://xn--273--84dlf.xn--plai/voprosy i otvety](http://xn--273--84dlf.xn--plai/voprosy_i_otvety).

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий [http://mschool.kubsu.ru](http://mschool.kubsu.ru;);
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины

6.1 Рекомендации по организации обучения

Освоение дисциплины «Методы искусственного интеллекта в задачах классификации» требует системного подхода и активной самостоятельной работы. По курсу предусмотрено проведение лекционных занятий, на которых дается систематизированный материал по дисциплине. В ходе лекций рассматриваются ключевые концепции. После каждой лекции рекомендуется выполнение практических заданий для закрепления ключевых понятий и методов и самостоятельная работа с дополнительным материалом и литературой.

Лабораторные занятия курса посвящены практическому освоению методов классификации. На занятиях студенты реализуют задачи классификации различных данных, в том числе в облачных средах, предоставленных партнерами.

При самостоятельной работе студентам необходимо изучать рекомендованную литературу в виде официальной документации к используемым открытым программным продуктам, облачным платформам.

6.2 Стратегии выполнения лабораторных работ

1. Подготовка данных:

- Освойте методы обработки данных (нормализация, кодирование категориальных признаков) на примере датасетов Iris, MNIST/CIFAR-10.
- Используйте pandas для анализа и scikit-learn для разделения выборки (train/test split).

2. Эксперименты с моделями:

- Начинайте с простых алгоритмов (логистическая регрессия), постепенно переходя к сложным (ансамбли, нейросети).
- Сравнивайте результаты по метрикам (F1-score, ROC-AUC) для разных моделей.

3. Анализ результатов:

- Визуализируйте матрицы ошибок, кривые обучения.
- Интерпретируйте работу моделей с помощью SHAP/LIME (Lab 9).
- Оптимизируйте гиперпараметры через GridSearchCV.

6.3 Проектная деятельность

Итоговая работа — решение сквозной задачи классификации (например, прогнозирование оттока клиентов, распознавание объектов на изображениях). Этапы:

1. Выбор датасета и постановка задачи.
2. Предобработка данных, feature engineering.
3. Обучение и валидация моделей (минимум 3 разных алгоритма).
4. Интерпретация результатов и формирование отчета.

Рекомендуется использовать [Kaggle](#) для поиска datasets и соревнований.

6.5 Рекомендации для студентов с ОВЗ

- Материалы предоставляются в адаптированных форматах: аудиоформат, электронные документы с увеличенным шрифтом.
- Консультации проводятся индивидуально (включая онлайн-формат).
- Лабораторные работы могут быть скорректированы (упрощенные датасеты, расширенные сроки сдачи).

Подход, определяющий установление соответствия кейсов ИП и УГТ (5-7), позволяет четко соотносить этапы развития технологии с вовлеченностью партнера и снижать риски при переходе от лабораторных испытаний к промышленному внедрению.

Кейсы ПАО «Сбербанк»

1. Генеративный ИИ для автоматического составления инвестиционных обзоров

Описание:

Аналитики Сбера ежедневно составляют десятки аналитических и инвестиционных обзоров по рынкам, компаниям, макроэкономике. Задача — исследовать применение LLM для генерации кратких сводок и аналитических отчетов на основе входных данных: биржевые котировки, макроэкономические показатели, рыночные события.

Цель:

Разработать инструмент, способный по структурированным данным и краткому описанию формировать инвестиционный обзор в деловом стиле.

Ожидаемый результат:

Модель, генерирующая аналитические тексты длиной 500–1000 слов с разделами «обзор событий», «рекомендации», «прогнозы», оформленные в формате банка.

2. НЛП-анализ жалоб клиентов в свободной форме

Описание:

В рамках клиентского сервиса Сбербанк обрабатывает обращения из чатов, мобильного приложения и жалобной формы. Требуется построить модель семантического анализа, выделяющую суть обращения, определяющую тональность и потенциальную серьезность инцидента.

Цель:

Автоматизировать классификацию обращений для ускорения маршрутизации и выявления повторяющихся болевых точек в продуктах и процессах.

Ожидаемый результат:

Прототип модели, автоматически выделяющей темы жалоб (например, «ошибка в приложении», «двойное списание»), их эмоциональную окраску и критичность.

3. Генерация сценариев фишинговых писем для обучения сотрудников**Описание:**

Банк проводит киберучения, включая рассылку тестовых фишинговых писем сотрудникам для повышения их устойчивости к социальным атакам. Проект предполагает использование генеративной модели для создания реалистичных фишинговых писем различных типов (поддельные счета, HR-запросы, ИТ-поддержка).

Цель:

Создать генератор, способный на основе заданных параметров (тема, стиль, уровень угрозы) создавать тексты фишинга для тренировок.

Ожидаемый результат:

Набор разнообразных примеров фишинга и оценка их эффективности по реакции сотрудников, а также классификация моделей угроз.

4. Генерация пользовательских сценариев работы в мобильном приложении**Описание:**

Банк хочет использовать генеративный ИИ для быстрой симуляции пользовательских сценариев — например, как клиент оформляет вклад, переводит средства, получает уведомление о риске мошенничества.

Цель:

Разработать генератор пошаговых сценариев пользовательского поведения с вариативностью (молодой клиент, пенсионер, ИП).

Ожидаемый результат:

Набор автоматически сгенерированных UX-сценариев, оформленных в виде сценариев для QA или UX-исследований, с логикой действий и типичными ошибками пользователя.

5. Генерация synthetic data для банковских моделей**Описание:**

Модели в Сбере требуют большого объема транзакционных и клиентских данных, которые нельзя использовать напрямую из-за требований ЦБ и ФЗ-152. Задача — разработать метод генерации синтетических банковских данных, максимально близких к реальным по распределениям и поведению.

Цель:

Создать безопасный pipeline генерации данных (например, транзакций, профилей клиентов, шаблонов расходов) для обучения моделей.

Ожидаемый результат:

Синтетический датасет и отчет о метриках приближенности к реальному (TSNE, K-L divergence и др.), с оценкой пригодности для обучения скоринговых или антифрод-моделей.

6. Модель анализа инвестиционной привлекательности малого бизнеса

Описание:

Банк активно развивает кредитование и инвестиционные инструменты для малого и среднего предпринимательства (МСП). Требуется создать модель, которая на основе открытых и банковских данных (выручка, расходы, тип деятельности, отзывы, онлайн-активность) оценивает инвестиционную привлекательность МСП.

Цель:

Разработать систему рейтинговой оценки компаний малого бизнеса с возможностью визуализации факторов и динамики показателей.

Ожидаемый результат:

Модель, присваивающая компании инвестиционный рейтинг (например, А–Е), объясняющая ключевые параметры и дающая рекомендации для инвестора.

7. Индивидуальная оценка кредитоспособности клиента на основе поведенческих данных

Описание:

Современный кредитный скоринг выходит за рамки финансовых данных. Необходимо исследовать, как поведенческие и цифровые следы (частота входа в мобильный банк, способы оплаты, география, время отклика) влияют на персональную оценку риска.

Цель:

Разработать ML-модель, оценивающую вероятность дефолта по нестандартным поведенческим признакам (возможно — с explainable AI).

Ожидаемый результат:

Прототип скоринговой модели, которая, помимо стандартных данных, учитывает цифровой профиль клиента и объясняет решения (SHAP, LIME и др.).

8. Предиктивная аналитика возврата инвестиций по инфраструктурным проектам

Описание:

В ряде случаев Сбербанк выступает участником/инвестором в региональных инфраструктурных проектах (жилые массивы, дороги, технопарки). Задача — оценить прогнозируемую эффективность вложений с учётом демографии, миграции, экономической активности.

Цель:

Разработать модель, прогнозирующую ROI на горизонте 3–5 лет, используя внешние источники данных: Росстат, ЕГРЮЛ, кадастр, соцмедиа.

Ожидаемый результат:

Аналитическая модель с возможностью геовизуализации и сценарного анализа (рост/спад, госпрограммы, смена трафика и т.п.).

Кейсы от «АВАЛАБ»

1. LLM и RAG для BI-системы Fastboard

Описание:

Для разрабатываемой компанией BI-системы Fastboard требуется разработать интерфейс на естественном языке для построения отчетов на больших массивах данных в ClickHouse. С помощью LLM необходимо классифицировать запросы пользователей на естественном языке и извлекать фактические параметры для дальнейшего вызова веб-сервиса отчетов.

Цель:

Разработать промпты для классификации и обработки запросов пользователей LLM и преобразования их к вызовам типовых отчетов с фактическими параметрами, извлекаемыми из запроса.

Ожидаемый результат:

Инструмент на основе LLM, позволяющий запрашивать данные о продажах.

2. Анализ обращений клиентов и CRM-переписки

Описание:

В службе клиентского сервиса застройщика ежедневно обрабатываются десятки обращений (e-mail, звонки, мессенджеры). Требуется реализовать систему семантического анализа и классификации NLU: выявлять суть обращений, уровень удовлетворенности, отслеживать повторяющиеся запросы.

Цель:

Автоматизировать первичный разбор и маршрутизацию запросов по тематике (сдача объекта, отделка, документы, жалоба и т.д.).

Ожидаемый результат:

Прототип, который выделяет суть обращений и формирует дашборд по текущим «болям» клиентов.

3. Модель прогнозирования сроков сдачи объектов на основе текстовых и визуальных данных

Описание:

Девелоперская компания ведёт аналитический архив по срокам строительства. С помощью мультимодальных моделей (текстовые отчёты + фото стройки) можно прогнозировать вероятность отклонения от графика сдачи.

Цель:

Разработать модель, которая по текущему статусу объекта (фото, отчёт СМР) оценивает риски задержек.

Ожидаемый результат:

Прототип, который показывает вероятность отклонений и даёт текстовые пояснения (основанные на распознанных признаках — «не завершены фасадные работы», «монтаж инженерии не начат»).

8. Обратная генерация — ИИ-помощник для покупателей квартир

Описание:

Будущие покупатели часто задают типовые вопросы о квартирах, планировках, ипотеке,

акциях, сроках. Вместо call-центра предлагается реализовать LLM-бота, который обрабатывает текстовые и голосовые запросы, показывает планировки, ссылается на PDF-документы и может «объяснять» информацию простым языком.

Цель:

Упростить коммуникацию с клиентами на этапе выбора квартиры и повысить качество первичного контакта.

Ожидаемый результат:

Демо-бот, способный отвечать на вопросы о жилом комплексе, ориентируясь в его характеристиках и маркетинговых документах.

7. Материально-техническое обеспечение по дисциплине (модулю)

1. Облачные платформы и сервисы

cloud.ru, YandexCloud, AWS/GCP/Azure – облачные вычисления

2. Системы управления версиями и коллаборации

Git/GitHub/GitLab – контроль версий кода и совместная разработка

3. Свободное ПО (Open Source)

GitLab, GIT, MLFlow, Docker, Kubernetes, Terraform.

Виртуальные машины, кластер Managed Kubernetes и ресурсы GPU в облаке предоставляется индустриальным партнером ПАО «Сбербанк»:

№	Продукт	Параметры продукта	Кол-во	Кол-во конфигураций	Ед. изм.
1	Виртуальная машина	Виртуальная машина 10% vCPU 2 vCPU 4 RAM	1	60	Шт
		ОС Ubuntu 22.04	1		Шт
		Системный диск SSD	1		Шт
			10		Гб
2	K8S	Master node 8 vCPU 16 RAM	1	1	Шт
		Worker node 10% доля 4 vCPU 32 RAM	5		Шт
		Worker node SSD-NVME	64		Гб
		Аренда публичного IP	1		Шт
3	ML Inference Instance Type GPU	Время работы в месяц	40	1	Ч
		Инстанс 8 x NVIDIA® H100 NVLink PCIe 160 vCPU 1520 GB RAM	1		Шт
		Количество запросов к ML-моделям	1		Млн. Шт
		Кэш ML-моделей	160		Гб
4	LLM	Токены GigaChat 2 Max	50		Млн. Шт

		Токены Embeddings	400		Млн. Шт
--	--	-------------------	-----	--	------------

Дополнительные облачные ресурсы предоставляются технологическим партнером Yandex Cloud.

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
1	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
2	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением
3	Практические занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
4	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
5	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
6	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.