

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
Б1.В.15 «Современные технологии машинного обучения»**

Направление подготовки 01.03.02 Прикладная математика и информатика

Направленность (профиль) Современные методы машинного обучения и
компьютерного зрения

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины Современные технологии машинного обучения составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 01.03.02 Прикладная математика и информатика

Программу составил(а):

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

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

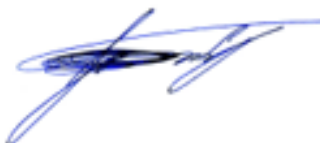


подпись

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

Цель освоения дисциплины «Современные технологии машинного обучения» является формирование у студентов систематизированных знаний, практических умений и навыков применения современных методов искусственного интеллекта, машинного обучения для решения задач анализа данных для дальнейшего применения в моделях различных предметных областей.

Дисциплина направлена на развитие навыков понимания основ методов машинного обучения, их реализации для различных предметных областей, и навыков оценки качества моделей. Студенты освоят полный цикл проекта — от предобработки данных и построения базовых моделей до использования продвинутых ансамблевых методов, анализа временных рядов и моделей с подкреплением, с акцентом на интерпретируемость и надежность решений.

1.2 Задачи дисциплины

- Сформировать понимание основных типов задач машинного обучения: обучение с учителем (классификация, регрессия), без учителя (кластеризация, снижение размерности) и с подкреплением.
- Изучить математические основы ключевых алгоритмов: линейные модели, деревья решений, методы ближайших соседей, ансамбли (bagging и boosting, включая XGBoost), вероятностные модели (Скрытые Марковские Модели).
- Освоить принципы оценки качества и валидации моделей, методы борьбы с переобучением и работы в условиях несбалансированных данных.
- Разобрать теорию работы Скрытых Марковских Моделей (НММ) и их применение для анализа последовательностей.
- Понять основы обучения с подкреплением (Reinforcement Learning), включая формализацию в виде Марковского процесса принятия решений (MDP).
- Развитие практических навыков использования методов машинного обучения в прикладных задачах. Например, обучение работать с разнотипными данными, визуализировать их, оценивать метрики качества работы алгоритмов.
- Ознакомление с примерами использования методов машинного обучения в реальных технологических задачах индустриальных партнеров.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Современные технологии машинного обучения» относится к дисциплинам по выбору, код Б1.В.15.

Дисциплина изучается в 4-м семестре. Для успешного освоения необходимы знания, полученные в дисциплинах: «Алгебра и введение в тензорный анализ», «Теория вероятностей и математическая статистика», «Многомерный статистический анализ», «Программирование».

Требованием к «входным» знаниям является понимание основ программирования на языке Python, базовых структур данных и алгоритмов, основ трансляции программ, основ математической логики.

Преподавание ведется в виде лекций и 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
- 2 Оптимизация производительности и масштабирование моделей
- 3 Разработка ML-пайплайнов и автоматизация процессов
- 4 Мониторинг качества моделей в продуктиве
- 5 Интеграция ML-решений с бизнес-приложениями

Роль 3: MLOps (Специалист по эксплуатации ИИ)

Задачи:

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

Роль 4: Software Engineer (Инженер-программист)

Задачи:

- 1 Проектирование и разработка веб, мобильных и десктоп-приложений на Python, Java, C/C++, JavaScript
- 2 Проектирование и разработка баз данных, веб-сервисов, чат-ботов
- 3 Интеграция информационных систем, программно-аппаратных комплексов
- 4 Математическое моделирование и разработка на основе моделей
- 5 *Командная разработка программного обеспечения*

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:

Соответствие **базовому уровню** освоения компетенций планиваемым результатам обучения и критериям их оценивания (оценка: **удовлетворительно /зачтено**):

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.1	<p>Различает основные типы задач машинного обучения и применяет на практике принципы их решения:</p> <p>Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением).</p> <p>Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn).</p> <p>Владеет навыками применения методов машинного обучения к данным из различных областей знания</p>
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	Знает и умеет обосновать способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Знает и умеет обосновать способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи.
ML-3.2	<p>Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ. Знает и умеет распознать и определить практическую задачу в терминах обучения на частично размеченных данных.</p> <p>Умеет применить стандартные алгоритмы для работы с данными и обучения моделей с использованием неполной разметки.</p>
ML-3.3	<p>Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами</p> <p>Умеет определить и сформулировать практическую задачу в терминах задачи ранжирования. Умеет применить простейшие методы ранжирования (pointwise подход)</p>
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей
ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Знает и Умеет применить стандартные библиотеки для обучения без учителя (ScikitLearn) понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE)
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач.

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

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.1	<p>Различает основные типы задач машинного обучения и применяет на практике принципы их решения:</p> <p>Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). Владеет навыками применения методов машинного обучения к данным из различных областей знания. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку</p>
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	<p>Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи</p> <p>Знает и Умеет обосновать способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Владеет навыками правильного выбора и обоснования методов решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку</p>
ML-3.2	<p>Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ. Умеет распознать и определить практическую задачу в терминах обучения на частично размеченных данных. Владеет навыками применения стандартных алгоритмов для работы с данными и обучения моделей с использованием неполной разметки. Владеет методами feature engineering: отбор создание и преобразование признаков.</p>
ML-3.3	<p>Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами</p> <p>Умеет определить и сформулировать практическую задачу в терминах задачи ранжирования. Умеет применить простейшие методы ранжирования (pointwise подход). Владеет навыками применения различных типов кросс-валидации, навыками оценки качества моделей с учетом bias-variance trade-off</p>
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей

ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Знает и умеет применить стандартные библиотеки для обучения без учителя (ScikitLearn), Знает и понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE). Владеет инструментами очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных.
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Знает алгоритмы обучения с подкреплением, Умеет обосновать способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач. Владеет навыками применения алгоритмы обучения с подкреплением.

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

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.1	Различает основные типы задач машинного обучения и применяет на практике принципы их решения: Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). Владеет навыками применения методов машинного обучения к данным из различных областей знания. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи Знает и умеет обосновать способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Умеет выбирать и обосновывать методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, Владеет навыками настройки базовых моделей и проводит их оценку

ML-3.2	Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ. Знает и умеет как распознать и определить практическую задачу в терминах обучения на частично размеченных данных. Умеет применить стандартные алгоритмы для работы с данными и обучения моделей с использованием неполной разметки. Владеет методами feature engineering: отбор создание и преобразование признаков.
ML-3.3	Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами Умеет определить и сформулировать практическую задачу в терминах задачи ранжирования. Владеет простейшими методами ранжирования (pointwise подход). Применяет различные типы кросс-валидации, Умеет оценить качество моделей с учетом bias-variance trade-off.
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей
ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Знает и умеет применить стандартные библиотеки для обучения без учителя (ScikitLearn) понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE). Владеет инструментами очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных. Умеет выбирать и настраивать алгоритмы кластеризации (DBSCAN, Mean Shift, Gaussian Mixture Models) и методы понижения размерности (UMAP, автоэнкодеры) в зависимости от специфики задачи. Интерпретирует полученные результаты и применяет их для обоснованных выводов
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач. ед. (144 часов), их распределение по видам работ представлено в таблице

Вид учебной работы	Всего часов	Семестры (часы)	
		4	
Контактная работа, в том числе:	68,3	68,3	
Аудиторные занятия (всего):	64	64	
Занятия лекционного типа	32	32	
Лабораторные занятия	32	32	
Занятия семинарского типа (семинары, практические занятия)			

Иная контактная работа:		4,3	4,3	
Контроль самостоятельной работы (КСР)		4	4	
Промежуточная аттестация (ИКР)		0,3	0,3	
Самостоятельная работа, в том числе:		40	40	
Курсовая работа				
Проработка учебного (теоретического) материала		15	15	
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)		15	15	
Реферат		-	-	
Подготовка к текущему контролю		10	10	
Контроль:		35,7	35,7	
Подготовка к экзамену		35,7	35,7	
Общая трудоемкость	час.	144	144	
	в том числе контактная работа	68,3	68,3	
	зач. ед	4	4	

2.2 Структура дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 4 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1	Раздел 1. Введение в машинное обучение и жизненный цикл проекта	10	4		2	4
2	Раздел 2. Классическое обучение с учителем	28	8		8	8
3	Раздел 3. Ансамбли и продвинутые методы	20	6		6	10
4	Раздел 4. Обучение без учителя	16	4		4	6
5	Раздел 5. Последовательности и временные ряды	30	6		4	6
6	Раздел 6. Обучение с подкреплением. Финальный проект	4	4		8	8
ИТОГО по разделам дисциплины		104	32		32	40
Контроль самостоятельной работы (КСР)		4				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к текущему контролю		35,7				
Общая трудоемкость по дисциплине		144				

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

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Введение в Machine Learning и жизненный цикл проекта	Что такое МЛ: задачи классификации, регрессии, кластеризации. Жизненный цикл Data Science проекта (CRISP-DM). Переобучение (Overfitting) и недообучение (Underfitting). Разбиение на выборки. Базовые метрики качества: Accuracy, Precision, Recall, F1-Score, MAE, MSE, R ² . Жизненный цикл проекта.	ЛР
2.	Предобработка данных и Feature Engineering	Работа с пропущенными значениями, категориальными признаками. Масштабирование признаков (StandardScaler, MinMaxScaler). Анализ и обработка выбросов. Создание новых признаков.	ЛР
3.	Линейные модели для регрессии и классификации	Линейная и логистическая регрессия. Градиентный спуск. Регуляризация: L1 (Lasso) и L2 (Ridge). Elastic Net.	ЛР
4.	Методы ближайших соседей и Наивный Байес	k-Nearest Neighbors (k-NN) для классификации и регрессии. Наивный байесовский классификатор (Gaussian, Multinomial). Преимущества, недостатки и сферы применения.	ЛР
5.	Деревья решений	Принцип работы дерева решений. Критерии (Gini, Entropy). Построение дерева, проблема переобучения и стрижка (pruning). Интерпретируемость деревьев.	ЛР
6.	Методы опорных векторов (SVM)	Принцип максимального зазора. Линейно разделимый и неразделимый случаи. Функции ядра (Kernel Trick) для нелинейных задач (RBF, полиномиальное). SVM для классификации и регрессии (SVR).	ЛР
7.	Ансамбли моделей: Bagging	Ансамблирование: бэггинг, случайный лес (Random Forest). Теория и преимущества бэггинга. Декорреляция деревьев. Out-of-Bag оценка (OOB). ExtraTrees.	ЛР
8.	Ансамбли моделей: Boosting. Введение в XGBoost	Принцип бустинга: AdaBoost, Gradient Boosting. XGBoost (eXtreme Gradient Boosting): архитектура, преимущества в скорости и точности. Ключевые гиперпараметры XGBoost.	ЛР
9.	Практическое применение XGBoost и интерпретация моделей	Стратегии тонкой настройки XGBoost. Методы интерпретации моделей: Feature Importance. Библиотека SHAP для объяснения предсказаний любой модели.	ЛР
10.	Кластеризация	Задачи кластеризации. Метрики расстояния. K-Means: алгоритм, инициализация, определение числа кластеров.	ЛР

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		Иерархическая кластеризация, DBSCAN и их сравнение.	
11.	Понижение размерности	PCA (Principal Component Analysis) для визуализации и сжатия признаков. LDA (Linear Discriminant Analysis) как метод с учителем. t-SNE и UMAP для нелинейной визуализации многомерных данных	ЛР
12.	Поиск аномалий и матричные разложения	Задача поиска аномалий (Anomaly Detection). Подходы: изолирующий лес (Isolation Forest), статистические методы.SVD (Singular Value Decomposition) и его применение.	ЛР
13.	Работа с временными рядами	Временные ряды: особенности, стационарность, автокорреляция. Классические методы: ARIMA, ETS (Exponential Smoothing).	ЛР
14.	Скрытые марковские модели (НММ)	Введение в марковские процессы.Теория СММ: скрытые состояния, наблюдаемые символы, матрицы переходов и эмиссий. Алгоритмы: Витерби (декодирование), Форвард-Бэквард (обучение). Библиотека hmmlearn в Python. Практические кейсы применения.	ЛР
15.	Введение в обучение с подкреплением (Reinforcement Learning). Финальный проект	Формализация задачи RL: агент, среда, состояние, действие, награда. Процесс принятия решений (Markov Decision Process). Q-learning и финальный проект курса.	ЛР

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.2 Занятия семинарского типа

Не предусмотрены

2.3.3 Лабораторные занятия

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Введение в Machine Learning и жизненный цикл проекта	Знакомство с окружением и базовый анализ данных Настройка окружения. Загрузка датасета (согласно варианту). EDA: info(), describe(), value_counts(), визуализация. Расчет базовых метрик "вручную".	ЛР

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
2.	Предобработка данных и Feature Engineering	Практика предобработки данных Обработка пропусков, кодирование категориальных переменных. Масштабирование числовых признаков. Разбиение на train/val/test.	ЛР
3.	Линейные модели для регрессии и классификации	Линейные модели Реализация линейной, Lasso, Ridge и логистической регрессии.	ЛР
4.	Методы ближайших соседей и Наивный Байес	k-NN и наивный Байес Реализация k-NN. Реализация Наивного Байеса. Сравнение моделей на валидационной выборке.	ЛР
5.	Деревья решений	Построение и визуализация дерева решений. Сравнение результатов с предыдущими линейными моделями.	ЛР
6.	Методы опорных векторов (SVM)	Методы опорных векторов (SVM) Реализация линейного SVM и SVM с RBF-ядром. Подбор гиперпараметра C и gamma. Визуализация разделяющей поверхности.	ЛР
7.	Ансамбли моделей: Bagging	Случайный лес Реализация Random Forest. Сравнение с одним деревом. Анализ важности признаков. ООБ-оценка.	ЛР
8.	Ансамбли моделей: Boosting. Введение в XGBoost	Знакомство с XGBoost Реализация XGBoost для классификации и регрессии. Подбор базовых гиперпараметров (n_estimators, max_depth, learning_rate).	ЛР
9.	Практическое применение XGBoost и интерпретация моделей	Тонкая настройка XGBoost и SHAP Использование GridSearchCV/RandomizedSearchCV для XGBoost. Анализ важности признаков и объяснение отдельных предсказаний с помощью SHAP.	ЛР
10.	Кластеризация	Кластеризация датасета Применение K-Means (с подбором K) и DBSCAN к датасету. Сравнение результатов кластеризации и их визуализация.	ЛР
11.	Понижение размерности	PCA и LDA Применение PCA для снижения размерности и визуализации. Применение LDA для того же датасета (с учителем). Сравнение.	ЛР
12.	Поиск аномалий и матричные разложения	Поиск аномалий Применение Isolation Forest и алгоритма DBSCAN для поиска аномалий. Визуализация и анализ выбросов.	ЛР
13.	Работа с временными рядами	Прогнозирование временных рядов (ARIMA) Загрузка и визуализация временного ряда. Проверка на стационарность, построение и оценка модели ARIMA.	ЛР

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
14.	Скрытые марковские модели (НММ)	Скрытые марковские модели с hmmlearn Моделирование простой последовательности (честная/нечестная монета). Обучение СММ и декодирование последовательности состояний алгоритмом Витерби. \ Практическое применение СММ. Применение СММ к реальным данным (например, прогнозирование режима работы устройства, простой анализ ДНК-последовательностей).	ЛР
15.	Введение в обучение с подкреплением (Reinforcement Learning). Финальный проект	Финальный проект Комплексный анализ нового датасета. Полный цикл: EDA, предобработка, построение нескольких моделей (например, Random Forest, XGBoost, SVM), их сравнение и интерпретация результатов.	РЗ

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.4 Примерная тематика курсовых работ (проектов)

Не предусмотрены

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

№	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 к текущему контролю	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры вычислительных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО

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 аудиофайла,
- 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 аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

- Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.
- 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.
- Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.
- 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.
 - Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

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

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

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

- 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.

Семестр	Вид занятия	Используемые интерактивные образовательные технологии	количество интерактивных часов
7	Л, ЛР, СРС	Практические занятия в режимах взаимодействия «преподаватель – студент» и «студент – студент»	64
Итого			64

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

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

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.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4. Оценочные и методические материалы

4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Современные технологии машинного обучения».

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	Раздел 1. Введение в машинное обучение и подготовка данных	ML-2; ML-3; ML-4; ML-6	Лабораторные работы №1-2	Вопросы к экзамену 1-7, 35
2	Раздел 2. Классическое обучение с учителем	ML-2; ML-3; ML-4; ML-6	Лабораторные работы №3-6	Вопросы к экзамену 8-14
3	Раздел 3. Ансамбли и продвинутое методы	ML-2; ML-3; ML-4; ML-6	Лабораторные работы №7-10	Вопросы к экзамену 15-20,34
4	Раздел 4. Обучение без учителя	ML-2; ML-3; ML-4; ML-6	Лабораторные работы №11-12	Вопросы к экзамену 21-26
5	Раздел 5. Последовательности и временные ряды	ML-2; ML-3; ML-4; ML-6	Лабораторные работы №13-14	Вопросы к экзамену 31-32
6	Раздел 6. Обучение с подкреплением. Финальный проект	ML-2; ML-3; ML-4; ML-6	Лабораторная работы №15 и финальный проект	Вопросы к экзамену 33-35

Показатели, критерии и шкала оценки сформированных компетенций

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

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценке качества моделей и работу с признаками
ML-2.1	Различает основные типы задач машинного обучения и применяет на практике принципы их решения: Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). Владеет навыками применения методов машинного обучения к данным из различных областей знания
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи.
ML-3.2	Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ.

	Распознает и определяет практическую задачу в терминах обучения на частично размеченных данных. Применяет стандартные алгоритмы для работы с данными и обучения моделей с использованием неполной разметки.
ML-3.3	Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами. Определяет и формулирует практическую задачу в терминах задачи ранжирования. Применяет простейшие методы ранжирования (pointwise подход)
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей
ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Применяет стандартные библиотеки для обучения без учителя (ScikitLearn) понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE)
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач.

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

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.1	Различает основные типы задач машинного обучения и применяет на практике принципы их решения: Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). Владеет навыками применения методов машинного обучения к данным из различных областей знания. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку
ML-3.2	Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ. Распознает и определяет практическую задачу в терминах обучения на частично размеченных данных. Применяет стандартные алгоритмы для

	работы с данными и обучения моделей с использованием неполной разметки. Владеет методами feature engineering: отбор создание и преобразование признаков.
ML-3.3	Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами Определяет и формулирует практическую задачу в терминах задачи ранжирования. Применяет простейшие методы ранжирования (pointwise подход). Применяет различные типы кросс-валидации Оценивает качество моделей с учетом bias-variance trade-off
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей
ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Применяет стандартные библиотеки для обучения без учителя (ScikitLearn) понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE). Владеет инструментами очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных.
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач.

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

ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.1	Различает основные типы задач машинного обучения и применяет на практике принципы их решения: Знает как различать основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Умеет применить типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). Владеет навыками применения методов машинного обучения к данным из различных областей знания. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку
ML-3	Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения
ML-3.1	Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Выбирает и обосновывает методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивает базовые модели и проводит их оценку
ML-3.2	Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ.

	Распознает и определяет практическую задачу в терминах обучения на частично размеченных данных. Применяет стандартные алгоритмы для работы с данными и обучения моделей с использованием неполной разметки. Владеет методами feature engineering: отбор создание и преобразование признаков.
ML-3.3	Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами. Определяет и формулирует практическую задачу в терминах задачи ранжирования. Применяет простейшие методы ранжирования (pointwise подход). Применяет различные типы кросс-валидации. Оценивает качество моделей с учетом bias-variance trade-off
ML-4	Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей
ML-4.1	Применяет алгоритмы кластеризации и понижения размерности для решения практических задач. Применяет стандартные библиотеки для обучения без учителя (ScikitLearn) понимает базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbSCAN иерархическая кластеризация) методы понижения размерности (PCA t-SNE). Владеет инструментами очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных. Выбирает и настраивает алгоритмы кластеризации (DBSCAN, Mean Shift, Gaussian Mixture Models) и методы понижения размерности (UMAP, автоэнкодеры) в зависимости от специфики задачи. Интерпретирует полученные результаты и применяет их для обоснованных выводов
ML-6	Способен применять алгоритмы обучения с подкреплением
ML-6.1	Обосновывает способы и варианты применения алгоритмов обучения с подкреплением в задачах ИИ, включая их преобразование и адаптацию к специфике задачи. Знает основные принципы обучения с подкреплением (агент, среда, награда) и умеет обосновать выбор простейших алгоритмов (Q-Learning, SARSA) для решения типовых задач.

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

***Практические задания по тематике лабораторных работ на основе кейсов
индустриальных партнеров***

Практическое задание 1: Классификация обращений клиентов строительной компании

Цель: Построить модель многоклассовой классификации текстовых обращений клиентов для автоматической маршрутизации.

Задачи:

1. Провести EDA датасета обращений (статистика по длине текстов, распределение тем, частотность слов)
2. Выполнить предобработку текстов: лемматизация, очистка от стоп-слов, обработка специальных терминов строительной тематики
3. Векторизовать тексты с помощью TF-IDF и Word2Vec
4. Обучить и сравнить модели:
 - Наивный Байес (MultinomialNB)
 - Logistic Regression

- Random Forest
 - XGBoost
5. Подобрать гиперпараметры с помощью GridSearch
 6. Проанализировать важность признаков с помощью SHAP
 7. Создать прототип пайплайна от сырого текста до предсказанной категории
- Метрики качества:** Accuracy, Precision/Recall по классам, F1-macro
Технологии: sklearn, pandas, nltk/spacy, xgboost, shap

Для решения задачи с подкреплением:

Возможное применение RL:

Динамическая маршрутизация обращений. Агент RL может учиться направлять обращения к определенным специалистам или в определенные отделы, чтобы минимизировать время решения проблемы или максимизировать удовлетворенность клиента.

Награда: Положительная обратная связь от клиента (например, оценка после решения проблемы) или уменьшение времени до закрытия обращения.

Метод: Можно использовать табличные методы (если количество состояний и действий невелико) или линейную аппроксимацию. Состояния могут быть представлены как категория проблемы, срочность, загрузка отделов. Действия — назначение на отдел или специалиста.

Практическое задание 2: Прогнозирование задержек сдачи объектов по операционным данным

Цель: Построить модель прогнозирования рисков срыва сроков сдачи строительных объектов.

Задачи:

1. Проанализировать временные ряды операционных данных (процент выполнения работ, поставки материалов, погодные условия)
2. Создать features из временных рядов: скользящие средние, тренды, лаги
3. Применить методы кластеризации (K-Means) для группировки объектов по схожести динамики строительства
4. Построить модель классификации (XGBoost) для прогнозирования срыва сроков
5. Использовать SMOTE для борьбы с дисбалансом классов
6. Интерпретировать модель с помощью SHAP - выявить ключевые факторы риска
7. Разработать дашборд с метриками рисков по объектам

Метрики качества: ROC-AUC, Precision-Recall curve, F1-score

Технологии: pandas, sklearn, xgboost, imbalanced-learn, plotly

Для решения задачи с подкреплением:

Возможное применение RL:

Оптимизация расписания строительных работ. Агент RL может учиться принимать решения о последовательности задач, распределении ресурсов, чтобы избежать задержек.

Награда: Снижение задержек (отрицательная награда за задержку, положительная за досрочное завершение этапов).

Метод: Из-за большого пространства состояний (прогресс по многим задачам, доступные ресурсы, погода) можно использовать линейную аппроксимацию или модельные методы, если мы можем построить модель динамики проекта.

Практическое задание 3: Классификация стадий строительства по отчетам подрядчиков

Цель: Автоматически определять текущую стадию строительства по текстовым отчетам прорабов.

Задачи:

1. Разметить данные: создать словарь строительных этапов (нулевой цикл, коробка, отделка и т.д.)
2. Применить тематическое моделирование (LDA) для выявления скрытых тем в отчетах
3. Построить модель многоклассовой классификации с помощью Random Forest и XGBoost
4. Реализовать ансамбль методов для улучшения качества:
 - Stacking классификаторов
 - Комбинирование текстовых и числовых features
5. Внедрить механизм обработки out-of-vocabulary слов через character-level TF-IDF
6. Создать систему валидации предсказаний на основе бизнес-логики (последовательность этапов)

Метрики качества: Weighted F1-score, Cohen's Kappa

Технологии: sklearn, gensim, xgboost, scipy

Для решения задачи с подкреплением:

Возможное применение RL:

Автоматизация сбора отчетов и уточняющих вопросов. Агент RL может учиться задавать уточняющие вопросы, если он не уверен в стадии, чтобы улучшить классификацию.

Награда: Увеличение точности классификации (но это отложенная награда, так как точность выяснится позже). Можно использовать внутренние награды за уверенность предсказания.

Метод: Табличные методы или линейная аппроксимация, где состояния — это текущий текст отчета и заданные вопросы, действия — задать вопрос или классифицировать.

Практическое задание 4: Поиск аномалий в операционной деятельности

Цель: Обнаруживать аномальные patterns в данных строительных проектов.

Задачи:

1. Проанализировать многомерные временные ряды (прогресс работ, расход материалов, трудозатраты)
2. Применить методы поиска аномалий:
 - Isolation Forest
 - Local Outlier Factor (LOF)
 - DBSCAN для временных рядов
3. Построить Скрытую Марковскую Модель (HMM) для нормального хода работ
4. Выявлять аномалии как отклонения от нормального состояния в HMM
5. Визуализировать обнаруженные аномалии на временной линии проекта
6. Разработать систему оповещений при обнаружении подозрительных patterns

Метрики качества: Precision@K, F1-score для аномалий

Технологии: sklearn, hmmlern, matplotlib, seaborn

Для решения задачи с подкреплением:

Возможное применение RL:

Активный поиск аномалий. Агент RL может учиться, какие данные проверять в первую очередь, чтобы быстрее находить аномалии. Например, в мониторинге сетевого трафика агент может выбирать, какие узлы проверять.

Награда: Обнаружение аномалии (положительная награда) и стоимость проверки (отрицательная награда).

Метод: Табличные методы, если узлов немного, или линейная аппроксимация для большого числа узлов.

Практическое задание 5: Сегментация клиентов по поведенческим паттернам

Цель: Кластеризовать клиентов для персонализации коммуникации.

Задачи:

1. Собрать features из истории взаимодействий (частота обращений, темы, длительность)
2. Применить методы снижения размерности:
 - PCA для визуализации
 - UMAP для нелинейного снижения размерности
3. Сравнить алгоритмы кластеризации:
 - K-Means (подобрать оптимальное K)
 - DBSCAN
 - Иерархическая кластеризация
4. Проинтерпретировать кластеры через анализ важных features
5. Построить модель классификации для автоматического определения сегмента новых клиентов
6. Разработать рекомендации по работе с каждым сегментом

Метрики качества: Silhouette Score, Calinski-Harabasz Index

Технологии: sklearn, umap, scipy, plotly

Для решения задачи с подкреплением:

Возможное применение RL:

Динамическая персонализация коммуникации. Агент RL может учиться, какие предложения делать клиенту из определенного сегмента, чтобы максимизировать конверсию или лояльность.

Награда: Клиент совершает целевое действие (покупка, отклик на предложение).

Метод: Поскольку сегментов может быть много и предложений много, можно использовать линейную аппроксимацию Q-функции, где признаки — это характеристики сегмента и история взаимодействий.

Практическое задание 6: Анализ тональности отзывов клиентов

Цель: Определить уровень удовлетворенности клиентов по текстовым отзывам.

Задачи:

1. Разметить датасет отзывов (позитивные/негативные/нейтральные)
2. Построить модель бинарной и многоклассовой классификации тональности
3. Применить разные подходы к векторизации:
 - TF-IDF с n-grams
 - Sentiment-specific features (словари эмоциональной лексики)
4. Сравнить производительность LinearSVM, Logistic Regression и XGBoost
5. Проанализировать наиболее значимые слова для каждого класса
6. Создать дашборд с динамикой удовлетворенности клиентов

Метрики качества: Accuracy, ROC-AUC, F1-score

Технологии: sklearn, xgboost, textblob, matplotlib

Для решения задачи с подкреплением:

Возможное применение RL:

Динамическое управление взаимодействием с клиентами. Например, если отзыв негативный, агент RL может решить, какую акцию предложить клиенту, чтобы вернуть его лояльность.

Награда: Изменение тональности следующего отзыва или повторная покупка.

Метод: Табличные методы или линейная аппроксимация.

Пример лабораторной работы *Корреляционный и регрессионный анализ данных*

1. С ресурса Всемирный Банк - World Development Indicators database для студентов сформирована CSV-таблица, содержащая информацию по 18 странам мира и по 22 характеристикам по каждой стране за годы (1990-2025).
2. Выполнить дескриптивный анализ и предварительную обработку данных. и Feature Engineering.
3. Необходимо выбрать данные по одной стране, согласно вашему варианту, построить кривую прироста ВВП, проанализировать взаимное влияние атрибутов, отобразить корреляцию:
 - a. Роста ВВП и прироста населения
 - b. Прироста населения на динамику безработицы
 - c. Изменения расходов на медицину и увеличения продолжительности жизни и смертность.
 - d. Прирост людей с высшим образованием на рост экспорта товаров и на прирост высокотехнологичного производства.
 - e. Расходов на образование на – кумулятивный прирост бакалавров среди женщин
 - f. Прирост людей с высшим образованием на развитие высоких технологий (прирост статей в научных журналах)
 - g. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.
4. Выполнить регрессионный анализ - прогноз по нескольким атрибутам.
5. Форма вашего датасета приведена в таблице:

Годы	G D P (V V P)	П ри ро ст В В П (G D P %)	Рост рождае мости, обусло вленны й квалиф икации й персона ла	Дина мика рост а рожд аемо сти на 1000 чело век	Заемщиков в банках на 1000 человек населения	. S L . U E M . A D V N . Z S	S L . U E M . B A S C . Z S	...	IP.J RN. AR TC. SC
	1	2	3	4	5	6	7	...	22

1990									
...									
2025									

Тестовые вопросы по дисциплине

1. **Какое разбиение данных НЕПРАВИЛЬНОЕ** для оценки качества модели?
 - a) Train/Test (70/30)
 - b) Train/Validation/Test (60/20/20)
 - c) Train/Test (50/50)
 - d) Train/Validation/Test (80/10/10)

 2. **При увеличении глубины дерева решений обычно:**
 - a) Увеличивается переобучение
 - b) Увеличивается недообучение
 - c) Снижается сложность модели
 - d) Увеличивается bias (смещение)

 3. **Метрика, наиболее подходящая для оценки модели в условиях сильного дисбаланса классов:**
 - a) Accuracy
 - b) Precision-Recall curve
 - c) R^2 score
 - d) Mean Absolute Error

 4. **One-Hot Encoding** применяется для:
 - a) Масштабирования числовых признаков
 - b) Кодирования порядковых категориальных признаков
 - c) Кодирования номинальных категориальных признаков
 - d) Обработки пропущенных значений

 5. **Градиентный спуск** используется для:
 - a) Визуализации многомерных данных
 - b) Оптимизации функции потерь
 - c) Кластеризации объектов
 - d) Уменьшения размерности данных
- Блок 2: Обучение с учителем (6-10)**
6. **L1-регуляризация (Lasso)** особенно полезна для:
 - a) Увеличения сложности модели
 - b) Отбора признаков (feature selection)
 - c) Уменьшения дисперсии модели
 - d) Работы с категориальными признаками

7. Алгоритм k-NN НЕЧУВСТВИТЕЛЕН к:

- a) Масштабированию признаков
- b) Выбору метрики расстояния
- c) Размеру обучающей выборки
- d) Порядку объектов в данных

8. Критерий Джини в деревьях решений измеряет:

- a) Линейную зависимость между признаками
- b) Степень неоднородности в узле
- c) Важность признаков
- d) Качество кластеризации

9. SVM с RBF-ядром позволяет:

- a) Строить только линейные разделяющие поверхности
- b) Эффективно работать с высокоразмерными данными
- c) Автоматически выбирать число кластеров
- d) Строить нелинейные разделяющие поверхности

10. Логистическая регрессия предсказывает:

- a) Числовое значение целевой переменной
- b) Вероятность принадлежности к классу
- c) Расстояние до ближайшего соседа
- d) Важность признаков

Блок 3: Ансамбли и XGBoost (11-14)

11. Основное отличие Random Forest от бэггинга над деревьями:

- a) Использование бустинга вместо бэггинга
- b) Случайный выбор признаков при разделении узлов
- c) Использование только одного дерева
- d) Отсутствие регуляризации

12. XGBoost отличается от обычного Gradient Boosting:

- a) Использованием только линейных моделей
- b) Регуляризацией и более эффективной реализацией
- c) Отсутствием возможности настройки гиперпараметров
- d) Только скоростью работы

13. Out-of-Bag оценка в Random Forest:

- a) Требуется отдельной тестовой выборки
- b) Использует объекты, не вошедшие в бутстрап-выборку
- c) Применима только для регрессии
- d) Всегда дает заниженную оценку качества

14. SHAP values используются для:

- a) Ускорения обучения модели
- b) Интерпретации предсказаний модели
- c) Уменьшения размерности данных
- d) Кодирования категориальных признаков

Блок 4: Обучение без учителя (15-18)

15. Метод локтя (Elbow method) применяется для:

- a) Определения оптимального числа кластеров в K-Means
- b) Выбора глубины дерева
- c) Настройки скорости обучения
- d) Определения числа главных компонент

16. Алгоритм DBSCAN особенно хорошо подходит для:

- a) Кластеризации сферических кластеров
- b) Выделения кластеров произвольной формы
- c) Работы только с малыми наборами данных
- d) Автоматического определения числа кластеров как $K=2$

17. PCA (Principal Component Analysis):

- a) Увеличивает размерность данных
- b) Сохраняет только дискретные признаки
- c) Уменьшает размерность, сохраняя максимальную дисперсию
- d) Только визуализирует исходные данные

18. Isolation Forest относится к алгоритмам:

- a) Кластеризации
- b) Понижения размерности
- c) Поиска аномалий
- d) Классификации

Блок 5: Временные ряды и НММ (19-22)

19. Стационарность временного ряда означает:

- a) Отсутствие тренда и сезонности
- b) Постоянство математического ожидания и дисперсии
- c) Линейную зависимость между наблюдениями
- d) Нормальное распределение ошибок

20. Скрытые Марковские модели (НММ) включают:

- a) Только наблюдаемые состояния
- b) Только скрытые состояния
- c) И скрытые, и наблюдаемые состояния
- d) Только матрицу переходов

21. Алгоритм Витерби используется для:

- a) Обучения НММ

- b) Нахождения наиболее вероятной последовательности скрытых состояний
- c) Кластеризации временных рядов
- d) Прогнозирования значений временного ряда

22. ARIMA модель включает компоненты:

- a) Авторегрессию, интегрирование, скользящее среднее
- b) Кластеризацию, классификацию, регрессию
- c) only скрытые состояния
- d) only наблюдаемые состояния

Блок 6: Обучение с подкреплением (23-25)

23. Markov Decision Process (MDP) включает:

- a) Только состояния и действия
- b) Состояния, действия, награды
- c) Только награды и действия
- d) Только состояния и награды

24. Q-learning является:

- a) Методом обучения с учителем
- b) Методом обучения без учителя
- c) Методом обучения с подкреплением
- d) Методом ансамблирования

25. Цель агента в обучении с подкреплением:

- a) Минимизировать функцию потерь
- b) Максимизировать совокупную награду
- c) Минимизировать число состояний
- d) Максимизировать число действий

Ключ для проверки:

1. c) Train/Test (50/50)
2. a) Увеличивается переобучение
3. b) Precision-Recall curve
4. c) Кодирования номинальных категориальных признаков
5. b) Оптимизации функции потерь
6. b) Отбора признаков (feature selection)
7. d) Порядку объектов в данных
8. b) Степень неоднородности в узле
9. d) Строить нелинейные разделяющие поверхности
10. b) Вероятность принадлежности к классу
11. b) Случайный выбор признаков при разделении узлов
12. b) Регуляризацией и более эффективной реализацией
13. b) Использует объекты, не вошедшие в бутстрап-выборку
14. b) Интерпретации предсказаний модели

15. а) Определения оптимального числа кластеров в K-Means
16. б) Выделения кластеров произвольной формы
17. с) Уменьшает размерность, сохраняя максимальную дисперсию
18. с) Поиска аномалий
19. б) Постоянство математического ожидания и дисперсии
20. с) И скрытые, и наблюдаемые состояния
21. б) Нахождения наиболее вероятной последовательности скрытых состояний
22. а) Авторегрессию, интегрирование, скользящее среднее
23. б) Состояния, действия, награды
24. с) Методом обучения с подкреплением
25. б) Максимизировать совокупную награду

Проверяемые компетенции комплексом практических заданий: ML-2.1; ML-3.1; ML-3.2; ML-3.3; ML-4.1; ML-6.1

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен)

Вопросы для подготовки к экзамену

Раздел 1: Основы машинного обучения и подготовка данных

1. Жизненный цикл проекта машинного обучения (CRISP-DM). Основные этапы и их содержание.
2. Проблемы переобучения (overfitting) и недообучения (underfitting). Методы их выявления и предотвращения.
3. Метрики качества для задач классификации: Accuracy, Precision, Recall, F1-Score, ROC-AUC. В каких случаях какие метрики предпочтительнее?
4. Метрики качества для задач регрессии: MAE, MSE, RMSE, R². Их интерпретация и особенности использования.
5. Методы обработки пропущенных значений в данных. Плюсы и минусы различных стратегий.
6. Кодирование категориальных признаков: One-Hot Encoding, Label Encoding, Target Encoding. Области применения и ограничения.
7. Методы масштабирования и нормализации признаков (StandardScaler, MinMaxScaler). Зачем они нужны и как влияют на разные алгоритмы?

Раздел 2: Обучение с учителем

8. Линейная регрессия. Метод наименьших квадратов. Предпосылки модели и их проверка.
9. Логистическая регрессия. Принцип работы, функция потерь. Интерпретация весов модели.
10. Регуляризация в линейных моделях: L1 (Lasso) и L2 (Ridge). Разница в их воздействии на веса модели и практическое применение.
11. Принцип работы алгоритма k-ближайших соседей (k-NN). Влияние гиперпараметра k и метрики расстояния на качество модели.
12. Деревья решений. Алгоритмы построения (ID3, CART). Критерии ветвления (энтропия, индекс Джини).
13. Гиперпараметры дерева решений (глубина, минимальное число samples в листе) и их влияние на переобучение.
14. Метод опорных векторов (SVM). Идея максимального зазора. Использование ядер (Kernel Trick) для нелинейного разделения.

Раздел 3: Ансамбли моделей

15. Бэггинг (Bootstrap Aggregating). Принцип, преимущества и почему он помогает бороться с переобучением.
16. Случайный лес (Random Forest). Отличия от бэггинга над деревьями. Понятие Out-of-Bag оценки.
17. Бустинг. Основная идея, отличия от бэггинга. Алгоритмы AdaBoost и Gradient Boosting.
18. XGBoost: ключевые особенности архитектуры, обеспечивающие скорость и точность. Основные гиперпараметры и стратегии их настройки.
19. Методы интерпретации моделей: Feature Importance. Разные типы важности признаков в деревьях и ансамблях.
20. Библиотека SHAP: принцип работы (теория Шепли) и виды графиков для интерпретации моделей.

Раздел 4: Обучение без учителя

21. Задачи кластеризации. Метрики расстояния/сходства в пространстве признаков.
22. Алгоритм K-Means: шаги работы, методы инициализации, способы определения оптимального числа кластеров.
23. Алгоритм DBSCAN: принцип работы, преимущества перед K-Means в задачах поиска аномалий и работы с кластерами произвольной формы.
24. Метод главных компонент (PCA). Геометрическая и алгебраическая интерпретация. Объясненная дисперсия.
25. Сравнение PCA и t-SNE: цели применения, интерпретация результатов, достоинства и недостатки.
26. Задача поиска аномалий (Anomaly Detection). Подходы на основе кластеризации (DBSCAN) и изолирующего леса (Isolation Forest).

Раздел 5: Последовательности и временные ряды

27. Скрытые Марковские модели (HMM). Основные элементы модели: скрытые состояния, наблюдаемые символы, матрицы переходов и эмиссий.
28. Алгоритмы для HMM: Витерби (нахождение наиболее вероятной последовательности состояний) и Баума-Велча (обучение модели).
29. Области применения Скрытых Марковских моделей в реальных задачах (например, анализ временных рядов, биоинформатика).
30. Классические методы прогнозирования временных рядов: ARIMA модель, её компоненты и требования к данным (стационарность).

Раздел 6: Обучение с подкреплением

31. Формальная постановка задачи обучения с подкреплением (Markov Decision Process). Основные компоненты: агент, среда, состояние, действие, награда.
32. Алгоритм Q-learning. Понятие Q-функции. Процесс обновления Q-значений. Роль гиперпараметров (скорость обучения, дисконт-фактор).

Общие и интеграционные вопросы

33. Построение полного пайплайна машинного обучения: от сырых данных до выбранной модели. Критерии выбора финальной модели для внедрения.

34. Сравнительный анализ рассмотренных алгоритмов (например, Logistic Regression vs. Random Forest vs. XGBoost). В каких ситуациях какой алгоритм предпочтительнее и почему?
35. Стратегия работы с данными, содержащими пропуски и выбросы, перед построением модели.

Перечень компетенций (части компетенции), проверяемых оценочным средством ML-2.1; ML-3.1; ML-3.2; ML-3.3; ML-4.1; ML-6.1

Практические задания к экзамену

1. Напишите базу знаний на языке Prolog или Python для модели родственных отношений. Реализуйте предикаты родитель/2, муж/1, жен/1, и определите на их основе предикаты отец/2, мать/2, брат/2, сестра/2.
2. Реализуйте простую производственную систему с механизмом прямого вывода. Создайте 5 правил и рабочую память. Продемонстрируйте, как система выбирает и применяет правила.
3. Постройте семантическую сеть для предметной области "Автомобили". Укажите узлы и связи (IS-A, HAS, PART-OF). Опишите графически или в текстовом виде.
4. Создайте фреймовую модель на основе семантической сети из задания 3. Представьте фреймы в виде структур с использованием слотов и значений.
5. Рассчитайте вероятность наступления события с помощью наивного Байесовского классификатора. Используйте простой набор данных (например: прогноз погоды — будет ли дождь?).
6. Напишите код для поиска кратчайшего пути с помощью алгоритма A*. Используйте граф с 6–8 узлами и заданными весами рёбер и эвристикой.
7. Создайте онтологию для предметной области "Птицы" с использованием RDF/OWL (в текстовом виде). Определите не менее 3 классов, 2 свойств и 2 подклассов.
8. Примените нечеткую логику для оценки "качества обслуживания" в ресторане. Определите лингвистические переменные ("низкое", "среднее", "высокое"), функции принадлежности и правило вывода.
9. Опишите простого рационального агента для задачи "поиск выхода из лабиринта". Опишите его перцепты, возможные действия и функцию рациональности.

Перечень компетенций (части компетенции), проверяемых оценочным средством ML-2.1; ML-3.1; ML-3.2; ML-3.3; ML-4.1; ML-6.1

Вопросы на анализ и сравнение

1. **Сравнение линейных моделей и деревьев решений**
Обсудите, в каких ситуациях линейные модели (например, линейная регрессия, логистическая регрессия) предпочтительнее деревьев решений, и наоборот. Какие особенности данных и задачи влияют на выбор между этими двумя подходами?
2. **Сравнение bagging и boosting**
В чем основные различия между bagging (например, Random Forest) и boosting (например, XGBoost)? Когда стоит использовать один метод, а когда другой?

3. **Выбор между k-NN и SVM**
Какие факторы следует учитывать при выборе между k-NN и SVM для задачи классификации? Как масштабирование признаков влияет на эти два алгоритма?
4. **Сравнение методов кластеризации: K-Means и DBSCAN**
В каких случаях K-Means будет работать лучше, чем DBSCAN, и наоборот? Как наличие шума и форма кластеров влияют на выбор алгоритма?
5. **Сравнение PCA и t-SNE**
Каковы основные цели использования PCA и t-SNE? Почему PCA часто используется для предобработки данных, а t-SNE — для визуализации?
6. **Выбор метода для задачи поиска аномалий**
Когда следует использовать Isolation Forest, а когда — DBSCAN для поиска аномалий? Какие ограничения у каждого из методов?
7. **Сравнение ARIMA и НММ для работы с временными рядами**
В чем разница между задачами, которые решают ARIMA и НММ? Когда уместно использовать ARIMA, а когда — НММ?
8. **Интерпретируемость моделей: линейные модели vs. случайный лес vs. XGBoost**
Какие из этих моделей легче интерпретировать и почему? Как такие методы, как SHAP, помогают в интерпретации сложных моделей?
9. **Влияние регуляризации на линейные модели и деревья**
Как регуляризация влияет на линейные модели и на ансамбли деревьев (например, в XGBoost)? Какие типы регуляризации используются в этих моделях?
10. **Сравнение методов обработки категориальных признаков**
Какие методы кодирования категориальных признаков вы знаете и в каких случаях каждый из них предпочтителен? Как наличие большого числа категорий влияет на выбор метода?
11. **Сравнение методов борьбы с дисбалансом классов**
Какие методы вы знаете для работы с дисбалансом классов? Как выбор метрики качества связан с дисбалансом? Когда стоит использовать oversampling, undersampling или взвешивание классов?
12. **Выбор стратегии обработки пропущенных значений**
Какие методы обработки пропущенных значений вы знаете? Как характер пропусков (MCAR, MAR, MNAR) влияет на выбор стратегии?
13. **Сравнение Q-learning и других методов обучения с подкреплением**
В чем преимущества и недостатки Q-learning по сравнению с другими алгоритмами обучения с подкреплением? Когда Q-learning может не справиться?
14. **Сравнение методов оптимизации в машинном обучении**
Чем градиентный спуск отличается от стохастического градиентного спуска? Какие методы оптимизации используются в XGBoost и в нейронных сетях?
15. **Сравнение методов оценки качества моделей**
Почему для оценки модели важно использовать отдельную тестовую выборку? В чем преимущества и недостатки кросс-валидации? Когда стоит использовать простое разделение на train/test, а когда кросс-валидацию?

Критерии оценки ответов:

Отличный ответ (5):

- Демонстрирует глубокое понимание принципов работы алгоритмов
- Приводит конкретные примеры из практики или исследований
- Анализирует trade-offs с учетом бизнес-контекста
- Предлагает обоснованные рекомендации для разных сценариев

Хороший ответ (4):

- Правильно сравнивает основные характеристики методов
- Объясняет преимущества и ограничения алгоритмов
- Приводит релевантные примеры применения
- Делает логичные выводы на основе сравнения

Удовлетворительный ответ (3):

- Называет основные различия между методами
- Описывает типичные случаи применения
- Допускает незначительные ошибки в деталях
- Недостаточно глубоко анализирует trade-offs

Неудовлетворительный ответ (2):

- Путаёт основные понятия и принципы
- Не может привести конкретные примеры
- Ответ поверхностный, без анализа
- Критические ошибки в понимании материала

Перечень компетенций (части компетенции), проверяемых оценочным средством ML-2.1; ML-3.1; ML-3.2; ML-3.3; ML-4.1; ML-6.1

4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Методические рекомендации, определяющие процедуры оценивания на экзамене:

Процедура промежуточной аттестации проходит в соответствии с Положением о текущем контроле и промежуточной аттестации обучающихся ФГБОУ ВО «КубГУ».

Итоговой формой контроля сформированности компетенций у обучающихся по дисциплине является экзамен. Студенты обязаны сдать экзамен в соответствии с расписанием и учебным планом.

ФОС промежуточной аттестации состоит из вопросов к экзамену и результатов текущего контроля.

Экзамен 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.

Форма проведения экзамена: устно.

Экзаменатору предоставляется право задавать студентам дополнительные вопросы по всей учебной программе дисциплины.

Результат сдачи экзамена заносится преподавателем в экзаменационную ведомость и зачетную книжку.

Оценивание уровня освоения дисциплины основывается на качестве выполнения студентом заданий текущего контроля и ответов на вопросы экзамена.

Критерии оценки:

1. Оценка ответов на экзаменационные вопросы (30% итоговой оценки)

Отлично (5)

Полные, развернутые ответы с демонстрацией глубокого понимания темы.
Использование примеров, формул, корректных терминов.
Умение анализировать и сравнивать методы (например, производственные системы vs. фреймы).
% выполнения: 90–100% (допускаются незначительные неточности).

Хорошо (4)

Ответы содержат основные идеи, но без углубленного анализа.
Возможны небольшие ошибки в деталях или формулировках.
% выполнения: 75–89%.

Удовлетворительно (3)

Ответы поверхностные, с существенными пробелами.
Отсутствие примеров или некорректное применение терминов.
% выполнения: 60–74%.

Неудовлетворительно (2)

Отсутствие понимания ключевых концепций.
Грубые ошибки или неспособность ответить на большую часть вопросов.
% выполнения: <60%.

2. Оценка выполнения проектов и лабораторных работ (60% итоговой оценки)

Отлично (5)

Полное выполнение всех этапов проектов и лабораторных работ, включая интеграцию LLM. Создание антагонистической игры с элементами обучения с подкреплением. Создание агента Robocode с обучением с подкреплением.
Четкая документация кода и анализ результатов.
% выполнения: 90–100%.

Хорошо (4)

Выполнены основные задачи лабораторных работ, но в проектах имеются существенные недочеты. Интеграция LLM не выполнена или работает в большинстве случаев некорректно.
% выполнения: 75–89%.

Удовлетворительно (3)

Решены базовые задачи лабораторных работ, проекты выполнены с критическими ошибками.
Низкое качество кода или отсутствие анализа.
% выполнения: 60–74%.

Неудовлетворительно (2)

Невыполнение ключевых этапов.
Код нерабочий или отсутствует.
% выполнения: <60%.

3. Оценка тестовых вопросов (10% итоговой оценки)

Отлично (5)

13–15 правильных ответов (87–100%).
Демонстрация уверенного владения терминологией и методами.

Хорошо (4)

10–12 правильных ответов (67–80%).
Незначительные ошибки.

Удовлетворительно (3)

7–9 правильных ответов (47–60%).
Путаница в базовых концепциях.

Неудовлетворительно (2)

Менее 7 правильных ответов (<47%).
Неспособность отличить ключевые модели представления знаний.

Итоговая оценка (суммарно)

Оценка	Экзамен (30%)	Практика (60%)	Тест (10%)	Общий %
Отлично (5)	90–100%	90–100%	87–100%	≥85%
Хорошо (4)	75–89%	75–89%	67–80%	70–84%
Удовлетворительно (3)	60–74%	60–74%	47–60%	55–69%
Неудовлетворительно (2)	<60%	<60%	<47%	<55%

Для допуска к экзамену необходимо выполнить **все лабораторные работы на минимум "удовлетворительно"**.

"Отлично" требует высоких результатов во всех компонентах (особенно в практических заданиях).

Практические кейсы оцениваются по:

- Корректности кода.
- Достижению метрик.
- Качеству отчета (анализ ошибок, визуализация).

Методические рекомендации, определяющие процедуры оценивания лабораторных работ:

Процедура оценивания лабораторных работ проходит в соответствии с Положением о текущем контроле и промежуточной аттестации обучающихся ФГБОУ ВО «КубГУ».

По каждой лабораторной работе оформляется отчет. Отчеты сдаются на проверку руководителю в течение курса по мере их выполнения, и защищаются студентами в установленном порядке.

При защите отчета студенту могут быть заданы вопросы и дополнительные задания по сути лабораторной работы, в том числе из списка контрольных вопросов к данной лабораторной работе. При неудовлетворительной оценке знаний студента по теме данного отчета, студент возвращается к повторному изучению соответствующих материалов, после чего допускается к повторной защите. Неудовлетворительно выполненный отчет также возвращается на доработку.

Отчет должен содержать заголовок, тему лабораторной работы, цель, задание, индивидуальную тему, описание хода выполнения работы, необходимые прикладные материалы (схемы, макеты документов и т.п.), в соответствии с требованиями к содержанию, и выводы по работе.

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 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.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

5.1 Основная литература:

1.1 Основная литература

(в том числе публикации конференций А*)

1. Митяков, Е. С. Искусственный интеллект и машинное обучение : учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2026. — 252 с. — ISBN 978-5-507-51198-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/507451> (дата обращения: 24.10.2025). — Режим доступа: для авториз. пользователей.
2. Баланов, А. Н. Машинное обучение и искусственный интеллект : учебное пособие для вузов / А. Н. Баланов. — 2-е изд., стер. — Санкт-Петербург : Лань, 2025. — 172 с. — ISBN 978-5-507-52891-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/462248> (дата обращения: 24.10.2025). — Режим доступа: для авториз. пользователей.
3. Машинное обучение : учебник : [16+] / Е. Ю. Бутырский, В. В. Цехановский, Н. А. Жукова [и др.]. — Москва : Директ-Медиа, 2023. — 368 с. : ил., табл., схем., граф. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=701807> (дата обращения: 24.10.2025). — Библиогр. в кн. — ISBN 978-5-4499-3778-0. — DOI 10.23681/701807. — Текст : электронный.
4. Биомедицинские сигналы и изображения в цифровом здравоохранении : хранение, обработка и анализ : учебное пособие / А. П. Немирко, Л. А. Манило, А. Ю. Долганов [и др.] ; под общ. ред. В. С. Кубланова ; Уральский федеральный университет им. первого Президента России Б. Н. Ельцина. — Екатеринбург : Издательство Уральского университета, 2020. — 243 с. : схем., табл. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=698902> (дата обращения: 24.10.2025). — Библиогр. в кн. — ISBN 978-5-7996-2990-8. — Текст : электронный.
2. Целых, А. Н. Применение временных рядов для анализа больших данных : учебное пособие по курсу «Математические методы анализа больших данных» : [16+] / А. Н. Целых, В. С. Васильев, Э. М. Котов ; Южный федеральный университет. — Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2021. — 86 с. : ил. — Режим доступа: по подписке. — URL:

- <https://biblioclub.ru/index.php?page=book&id=691448> (дата обращения: 24.10.2025). – Библиогр. в кн. – ISBN 978-5-9275-3983-3. – Текст : электронный.
3. Целых, А. Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» : [16+] / А. Н. Целых, А. А. Целых, Э. М. Котов ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2021. – 130 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=683920> (дата обращения: 24.10.2025). – Библиогр. в кн. – ISBN 978-5-9275-3783-9. – Текст : электронный.
 4. A Survey on Statistical Theory of Deep Learning: Approximation, Training Dynamics, and Generative Models. Namjoon Suh and Guang Cheng. *Department of Statistics and Data Science, UCLA, 2024.*
 5. Deception detection with machine learning: A systematic review and statistical analysis Alex Sebastião Constâncio, D. Tsunoda, D. Carvalho Published in PLoS ONE 9 February 2023.
 6. 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
 7. 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
 8. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." *Oncotarget* 8.7 (2016): 10883.
 9. 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.
 10. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." *Frontiers in pharmacology* 11 (2020): 565644.
 11. 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.
 12. 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.
 13. 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.
 14. 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

5.2 Дополнительная литература:

1. Разметка данных в машинном обучении: процесс, разновидности и рекомендации [Электронный ресурс]. - URL: <https://habr.com/ru/articles/678524/>. - (Дата обращения: 10.10.2025).
2. Неструктурированные данные: примеры, инструменты, методики и

- рекомендации [Электронный ресурс]. - URL: <https://habr.com/ru/articles/756454/>. - (Дата обращения: 10.10.2025).
3. Structured vs. Unstructured Data: What's the Difference? [Электронный ресурс]. - URL: <https://www.coursera.org/articles/structured-vs-unstructured-data>. - (Дата обращения: 10.10.2025).
 4. What is unstructured data? [Электронный ресурс]. - URL: <https://www.elastic.co/what-is/unstructured-data>. - (Дата обращения: 10.10.2025).
Kovriguina, L., Shilin, I., Putintseva, A., Shipilo, A. Multilevel Annotation in the
 5. Corpus for Parsing Russian Spontaneous Speech. In: Karpov, A., Jokisch, O., Potapova, R. (eds) Speech and Computer. SPECOM 2018. Lecture Notes in Computer Science(), vol 11096. Springer, 2018 - 311-320 p.
 6. Anthony S. Training Data for Machine Learning. O'Reilly Media, 2023. - 332 p. books on Data Annotation [Электронный ресурс]. - URL: https://www.aistartups.org/books/data_annotation/. - (Дата обращения: 01.10.2025).

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/>
9. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
10. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

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

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>

Бесплатные образовательные ресурсы

1. Jupyter Notebook – интерактивные вычисления
2. Visual Studio Code – редактор кода с поддержкой Python
3. Google Scholar/arXiv – доступ к научным публикациям

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.uceba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.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;>
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 (модуля)

По курсу предусмотрено проведение лекционных занятий, на которых дается систематизированный материал по основным направлениям искусственного интеллекта. В

ходе лекций рассматриваются ключевые концепции: от представления знаний (семантические сети, фреймы, онтологии) до современных подходов, таких как большие языковые модели, интеллектуальные агенты и системы управления знаниями. Особое внимание уделяется практическим аспектам: разбираются примеры применения ИИ в реальных задачах (экспертные системы, автоматическое планирование), обсуждаются типичные проблемы и ограничения при внедрении ИИ-решений. После каждой лекции рекомендуется выполнение практических заданий для закрепления ключевых понятий и методов.

Лабораторные занятия курса посвящены практическому освоению методов построения систем искусственного интеллекта. На занятиях студенты реализуют и анализируют модели представления знаний, работают с продукционными системами, фреймами и семантическими сетями, а также исследуют работу алгоритмов поиска (A^* , BFS, DFS), работают с нечеткой логикой. Используются реальные сценарии, такие как построение онтологий, моделирование агентов, реализация простых экспертных систем. Студенты применяют такие языки, как Prolog, RDF/OWL, Java, Python. Используют знания, полученные в других курсах, для интеграции больших языковых моделей с ИИ на базе структурированных знаний. После каждого лабораторного занятия предлагаются задания для самостоятельного выполнения — от построения семантических сетей до применения алгоритмов вывода.

При самостоятельной работе студентам необходимо изучать рекомендованную литературу (учебники, научные статьи, документацию библиотек) для глубокого понимания теоретических основ и современных подходов в области искусственного интеллекта. Выполняя проектные задания, студент должен уметь: формулировать задачу (например, моделирование знаний, поиск решения, построение агента); подбирать и адаптировать подходящие методы (от классических логических систем до вероятностных и нейросетевых подходов); реализовывать и оценивать эффективность выбранных решений. Особое внимание уделяется навыкам анализа и сравнения методов, а также критическому осмыслению их применимости в конкретной предметной области.

Важнейшим компонентом курса является самостоятельная проектная работа, в ходе которой студент разрабатывает законченное решение на основе ИИ-технологий — например, антагонистическую игру, интеллектуального агента, систему на основе нечеткой логики. Такой проект позволяет закрепить навыки проектирования, реализации и оценки систем искусственного интеллекта, а также научиться интегрировать разные методы и подходы в единое решение.

Для студентов с ограниченными возможностями здоровья предусмотрены дополнительные индивидуальные консультации, на которых преподаватель подробно разъясняет сложные аспекты дисциплины, помогает адаптировать практические задания и обеспечивает специальные условия для освоения методов работы с системами искусственного интеллекта. Индивидуальный подход позволяет таким студентам полноценно участвовать в учебном процессе и достигать требуемых результатов обучения.

6.3. 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. NLP-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 серьёзность 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. Мультиmodalный ассистент для банковских отделений

Описание:

Физические отделения Сбербанка внедряют интерактивных консультантов. Предполагается создание мультиmodalного ИИ-ассистента, который воспринимает речь и визуально ориентируется в пространстве (распознаёт клиента, документы, банкоматы), а также отвечает голосом.

Цель:

Разработать базовый прототип, имитирующий функциональность помощника: ответы на типовые запросы, визуальные подсказки, навигация по отделению.

Ожидаемый результат:

Интерактивная модель, объединяющая голосовой ввод, зрительное восприятие (например, QR-код паспорта), текстовый вывод и жестовую реакцию.

5. Объяснимость и контроль генеративных моделей в банковском ИИ**Описание:**

Банк активно использует LLM и NLP-сервисы (в чат-ботах, генерации шаблонов ответов, автоответах на e-mail), однако встает вопрос: как объяснять и контролировать поведение таких моделей, особенно в юридически значимых коммуникациях?

Цель:

Исследовать подходы к трассировке решений LLM (например, через логирование reasoning chain, пост-фильтрацию ответов, встроенные правила).

Ожидаемый результат:

Концепция системы explainability + compliance-модуля, обеспечивающего соответствие генерации стандартам банка и регулятора.

6. 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.

7. 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.

8. 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.

9. 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.).

10. 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.).

11. Анализ поведения пользователей в экосистеме цифрового рубля**Описание:**

Сбербанк участвует в пилотных проектах по внедрению цифрового рубля. Интерес представляет исследование пользовательских паттернов: как изменяются модели потребления, скорости операций, уровень доверия, сравнение с классическим безналом.

Цель:

Построить модель анализа поведения клиентов, участвующих в транзакциях с цифровым рублем: частота, средний чек, контексты.

Ожидаемый результат:

Отчёт и ML-модель, классифицирующая типы пользователей и выявляющая ключевые различия в предпочтениях и барьерах цифровой валюты.

12. Сравнение text2video / text2img моделей**Описание:**

Сбербанк заинтересован в сравнении text2video / text2img моделей (открытые модели, особенно китайские). Задача требует применения облачных ресурсов партнера для машинного обучения. От студентов требуется навык запуска открытых моделей, планирования, структурирования и логирования экспериментов, совместной работы. Задача может быть распараллелена для сравнения множества моделей независимо в группе студентов.

Цель:

Провести сравнение работы актуальных открытых моделей text2video / text2img.

Ожидаемый результат:

Таблица с результатами экспериментов модель / репозиторий / функционал / требования / оценка производительности / X примеров генераций (было/стало), human_eval по принципу арены (какая лучше)

Кейсы от «АВАЛАБ»**1. LLM и RAG для BI-системы Fastboard**

Описание:

Для разрабатываемой компанией ВІ-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. Генеративный ИИ для создания проектной документации по ТЗ**Описание:**

В рамках проектирования объектов девелоперской компании архитекторы и инженеры тратят значительное время на подготовку текстовой проектной документации (обоснование решений, пояснительные записки, описания инженерных систем). Задача — исследовать возможность использования LLM для генерации черновиков проектной документации на основе исходных данных: этажность, материалы, климат, назначение, нормы.

Цель:

Разработать прототип текстового генератора, который помогает специалистам быстрее формировать документацию в соответствии с шаблонами и нормативами.

Ожидаемый результат:

Инструмент на основе LLM, создающий логически стройный и нормативно грамотный текст, поддающийся быстрой редакции инженером.

4. Мультимодальный агент для анализа строительных площадок**Описание:**

ООО «АВА ЛАБ» разрабатывает систему для мониторинга строительных объектов. Требуется создать прототип мультимодального ИИ-агента, способного анализировать изображения со стройплощадки (видео/фото), а также принимать голосовые и текстовые запросы (например, «проверь монтаж перекрытия на 5 этаже»).

Цель:

Объединить возможности компьютерного зрения (распознавание стадии строительства, техники, нарушений) и НЛП (понимание запросов, отчетов).

Ожидаемый результат:

Интерактивный агент, который на запрос специалиста может показать нужный участок, прокомментировать прогресс, зафиксировать нарушения.

4. Генерация рекламного контента для жилых комплексов**Описание:**

«АВА ГРУПП» регулярно запускает маркетинговые кампании для жилых комплексов. Необходимо исследовать использование диффузионных моделей для генерации

изображений (визуализации интерьеров, окрестностей, видов из окон) и LLM — для описаний квартир, преимуществ района, инфраструктуры.

Цель:

Создать инструменты для быстрой генерации продающих материалов без привлечения дизайнеров и копирайтеров на первых этапах.

Ожидаемый результат:

Набор сгенерированных карточек объектов с текстом, изображением и логикой «живого» рекламного сообщения.

6. Генерация документации и шаблонов договоров

Описание:

Юридический департамент регулярно работает с договорами долевого участия, актами приёма-передачи и другими документами. Использование LLM может значительно сократить время на подготовку черновиков — достаточно ввести параметры сделки.

Цель:

Создать систему, которая генерирует адаптированные тексты документов по вводным данным (тип объекта, этаж, площадь, ФИО, сроки и пр.).

Ожидаемый результат:

Генератор документов в формате Word или PDF с автоматической подстановкой параметров и соблюдением юридического стиля.

7. 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.

КЕЙСЫ ДЛЯ ООО «СвязьРесурс-Кубань»

Описание:

Компания ООО "СвязьРесурс-Кубань" оказывает услуги связи. Работа с клиентами автоматизирована на базе CRM Битрикс 24. Для компании актуальны вопросы разработки первоначальных версий документов с помощью LLM и в перспективе автоматизации генерации большого количества документов по шаблонам с помощью LLM и RAG системы с интеграцией с Битрикс 24. Задачи включают в себя:

1. Разработка библиотеки промптов для генерации регламентов описания бизнес-процессов Битрикс 24.
2. Разработка библиотеки промптов для генерации техзаданий на основе параметров оказания услуг.
3. Разработка библиотеки промптов для генерации коммерческих предложений на основе параметров оказания услуг.
4. Разработка библиотеки промптов для генерации скриптов работы технической поддержки.
5. Разработка библиотеки промптов для генерации скриптов работы отдела продаж.
6. Апробация и сравнение различных языковых моделей для решения задач.

Цель:

Автоматизировать работу сотрудников по составлению типовых документов.

Ожидаемый результат:

Библиотека промптов и рекомендации по использованию LLM для решения поставленных задач.

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
		Аренда публичного IP	1		Шт
2	Виртуальная машина с GPU	Виртуальная машина с GPU NVIDIA® Tesla® V100 2 GPU 8 vCPU 128 ГБ RAM	1	1	Шт
		ОС Ubuntu_24.04	1		Шт
		Системный диск SSD	1		Шт
			2000		Гб
		Диск SSD	1		Шт
			4096		Гб
		Диск SSD	1		Шт
			4096		Гб
	Аренда публичного IP	1	Шт		
3	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
4	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
5	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.