

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

**Рабочая программа дисциплины (модуля)
«Рекомендательные системы»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Backend-разработка

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	4
3. Тематический план.....	6
4. Содержание дисциплины (модуля).....	6
5. Учебно-методическое обеспечение	7
6. Материально-техническое обеспечение	7
7. Методические и оценочные материалы	9

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Рекомендательные системы» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 02.04.01 Математика и компьютерные науки, профиль Backend-разработка, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Рекомендательные системы» важно для понимания методов персонализации контента и улучшения пользовательского опыта в различных цифровых сервисах. Эти знания позволяют создавать эффективные алгоритмы, которые помогают бизнесам увеличивать вовлеченность клиентов и оптимизировать процессы принятия решений.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Backend-разработка и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре, доступна для прохождения при условии успешного завершения дисциплины (модуля) «Machine Learning (Машинное обучение)».

Цель изучения дисциплины (модуля): освоение методов и алгоритмов, которые позволяют создавать персонализированные рекомендации для пользователей, улучшая их опыт взаимодействия с продуктами и услугами.

Задачи изучения дисциплины (модуля):

— формирование знаний и развитие понимания по темам: польза и финансовый эффект рекомендательных платформ, устройство современных индустриальных рекомендательных систем, способы повышения объяснимости и уменьшения смещений рекомендательных алгоритмов;

— освоение умений проводить разведывательный анализ рекомендательных данных, реализовывать классические рекомендательные алгоритмы, собирать рекомендательные пайплайны из базовых алгоритмов;

— формирование навыка владения современными библиотеками для построения рекомендательных систем, методами оценки качества рекомендательных алгоритмов, методами работы с матричными, тензорными, последовательными и графовыми данными.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-2.	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1.	Знает основные математические модели и методы, используемые в естественных науках, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных
		ОПК-2.2	Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в естественных науках, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований
		ОПК-2.3	Имеет практический опыт создания и исследования математических моделей в рамках научных проектов или

			исследований, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы
ПК-3.	Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата	ПК-3.1.	Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта
		ПК-3.2.	Умеет применять математические и компьютерные методы для решения конкретных задач, формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений
		ПК-3.3.	Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений
ПК-4.	Способен публично представлять собственные и известные научные результаты	ПК-4.1.	Знает основные принципы эффективного публичного выступления, методы визуализации данных и основные требования к научным презентациям, включая структуру и содержание
		ПК-4.2.	Умеет четко и логично формулировать свои научные результаты, адаптируя их для различных аудиторий, а также использовать визуальные средства для улучшения восприятия информации
		ПК-4.3.	Имеет практический опыт участия в научных конференциях, семинарах или других мероприятиях, где успешно представлял свои и известные научные результаты, получая обратную связь и взаимодействуя с аудиторией

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Аудиторная работа		Контроль	Самостояте льная работа	
Лекции	Семинары (практичес кие занятия)					
1	Приложения и оценка рекомендательных систем	3	3		16	Домашние задания
2	Классические алгоритмы	3	3		16	Домашние задания, Тест
3	Многостадийные системы	3	3		16	Домашние задания
4	Проблемы и смещения	3	3		16	Домашние задания, Тест
5	Индустриальные технологии	3	3		16	Домашние задания, Соревнование
	<i>Зачет с оценкой</i>			4		
	<i>Итого:</i>	<i>15</i>	<i>15</i>	<i>4</i>	<i>80</i>	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	<i>114</i>				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	<i>3</i>				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Приложения и оценка рекомендательных систем	Бизнес-приложения. Оценка качества.
2	Классические алгоритмы	Подходы на базе подобия. Модели на графах. Матричная факторизация. Продвинутая факторизация. Модели на последовательностях.
3	Многостадийные системы	Нейросетевое ранжирование.
4	Проблемы и смещения	Объяснимость и интерпретируемость. Проблемы и смещения. Обучение с подкреплением.
5	Индустриальные технологии	Доменные специфики. Обсуждение практических вопросов.

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Константинова, Е. В. Теория графов: алгебраическая теория : учебник для вузов / Е. В. Константинова. — Москва : Издательство Юрайт, 2025. — 123 с. — (Высшее образование). — ISBN 978-5-534-20172-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/569211>.

Дополнительная литература:

1. Скорубский, В. И. Математическая логика : учебник и практикум для вузов / В. И. Скорубский, В. И. Поляков, А. Г. Зыков. — Москва : Издательство Юрайт, 2025. — 211 с. — (Высшее образование). — ISBN 978-5-534-01114-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561076>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое

Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Рекомендательные системы» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, практические занятия, домашние задания, тесты, соревнование, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (практическом занятии) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины (модуля).

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины (модуля), понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания студентами важных методологических категорий.

Соревнование – организованное мероприятие, в рамках которого участники соперничают друг с другом для достижения определенной цели, демонстрируя свои навыки, знания или способности в заданной области.

В процессе подготовки к соревнованию опирайтесь на следующие рекомендации:

1. **Понимание задачи:** внимательно изучите условия соревнования и четко определите задачу, которую необходимо решить. Убедитесь, что вы понимаете, какие метрики будут использоваться для оценки ваших результатов.

2. **Сбор данных:** ознакомьтесь с предоставленным набором данных. Проведите анализ данных, выявите пропущенные значения, выбросы и другие особенности, которые могут повлиять на модель.

3. **Выбор алгоритмов:** исследуйте различные алгоритмы машинного обучения, подходящие для вашей задачи. Начните с простых моделей, затем переходите к более сложным, если это необходимо.

4. **Обучение и валидация:** разделите данные на обучающую и тестовую выборки. Используйте кросс-валидацию для оценки качества модели и избежания переобучения.

5. **Оптимизация гиперпараметров:** Экспериментируйте с настройками алгоритмов, для нахождения оптимальных гиперпараметров.

6. **Документация и презентация:** ведите записи о своих подходах, результатах и выводах. Подготовьте ясную и структурированную презентацию для финального отчета.

7. **Обратная связь и улучшение:** после получения результатов соревнования проанализируйте ошибки и недостатки вашей модели. Используйте этот опыт для улучшения своих навыков в будущем.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Рекомендательные системы»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы,
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
3	Не сдан	Не зачтено	Студент не овладел обязательным

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
2	Не сдан	Не зачтено	минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Рекомендательные системы» оценивается следующим образом:

Активность	Вес	Количество	Описание
<i>Накопительная оценка</i>			
Домашние задания	60%	3	Набор задач по темам недели
Тесты		2	Письменная работа с набором задач, которые нужно решить за ограниченное время
Соревнование		1	Kaggle-style соревнование с задачей на RecSys
<i>Промежуточная аттестация</i>			
Зачет с оценкой	40%	1	Мок-интервью — имитация в формате реального собеседования по ML System Design рекомендательных систем

Формула расчёта итоговой оценки по дисциплине (модулю) «Рекомендательные системы»: « $0,6 \times$ накопительная оценка ($0,5 \times$ среднее за домашние задания + $0,2 \times$ среднее за тесты + $0,3 \times$ соревнование) + $0,4 \times$ зачет с оценкой».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

Задание 1.

Разработайте концепцию бизнес-приложения, которое решает конкретную проблему в вашей целевой отрасли (например, здравоохранение, образование или электронная коммерция). Опишите ключевые функции приложения, целевую аудиторию и предполагаемые метрики успеха. Создайте прототип интерфейса пользователя с помощью любого инструмента для проектирования (например, Figma или Adobe XD).

Задание 2.

Выберите три различных метода оценки качества программного обеспечения (например, тестирование, ревью кода и автоматизированные тесты) и проведите сравнительный анализ их преимуществ и недостатков. Напишите отчет, в котором объясните, как каждый из методов может быть применен в процессе разработки, и предложите рекомендации по их комбинированию для достижения наилучших результатов.

Задание 3.

Исследуйте два подхода на базе подобию (например, коллаборативная фильтрация и контентная фильтрация) и создайте модель рекомендательной системы, используя один из этих подходов. Затем постройте графовую модель, представляющую пользователей и товары, и проанализируйте, как графовые алгоритмы (например, PageRank или алгоритмы поиска по графам) могут улучшить качество рекомендаций. Подготовьте презентацию, в которой представите результаты вашего анализа и модели.

Домашнее задание 2.

Задание 1.

Изучите основы матричной факторизации и реализуйте алгоритм, который использует этот метод для решения задачи рекомендательной системы. Используйте набор данных о предпочтениях пользователей (например, MovieLens). В вашем проекте:

- Реализуйте матричную факторизацию с использованием градиентного спуска.
- Оцените качество рекомендаций с помощью метрик, таких как RMSE или MAE.
- Подготовьте отчет, в котором опишите процесс, результаты и возможные улучшения.

Задание 2.

Исследуйте методы продвинутой факторизации, такие как факторизация через скрытые факторы и регуляризация. Реализуйте один из этих методов на том же наборе данных, что и в первом задании, добавив:

- Регуляризацию для предотвращения переобучения.
- Сравнение производительности с обычной матричной факторизацией.
- Подготовьте презентацию, в которой представите результаты и обсудите, как продвинутая факторизация влияет на качество рекомендаций.

Задание 3.

Создайте нейросетевую модель для задачи ранжирования, используя архитектуру, такую как RankNet или LambdaRank. В вашем проекте:

- Выберите набор данных, содержащий информацию о пользователях и их предпочтениях (например, набор данных о поисковых запросах).
- Обучите модель и оцените качество ранжирования с использованием метрик, таких как NDCG или MAP.
- Подготовьте отчет, в котором опишите архитектуру модели, результаты и предложите возможные улучшения для повышения качества ранжирования.

Домашнее задание 3.

Задание 1.

Изучите методы объяснимости и интерпретируемости моделей машинного обучения, такие как LIME и SHAP. Реализуйте один из этих методов на уже обученной модели (например, на модели классификации). В вашем проекте:

- Выберите набор данных (например, о кредитных заявках или медицинских диагнозах).
- Объясните предсказания модели для нескольких примеров, используя выбранный метод.
- Подготовьте отчет, в котором опишите, как интерпретируемость влияет на доверие пользователей к модели и какие выводы можно сделать из полученных объяснений.

Задание 2.

Исследуйте различные типы смещений (bias) в данных и их влияние на модели машинного обучения. Выберите один из наборов данных, в котором присутствуют смещения, и выполните следующие шаги:

- Проанализируйте данные на наличие смещений (например, смещение выборки или смещение подтверждения).
- Обучите модель на исходных данных и оцените ее производительность.

- Попробуйте устранить или уменьшить смещения (например, с помощью методов переподборки или взвешивания) и сравните результаты. Подготовьте отчет с анализом и выводами.

Задание 3.

Исследуйте методы обучения с подкреплением и выберите конкретную доменную специфику (например, игры, робототехника или управление ресурсами). В вашем проекте:

- Реализуйте простую модель обучения с подкреплением (например, Q-learning или DQN) для решения задачи в выбранной доменной специфике.
- Оцените эффективность модели, используя соответствующие метрики (например, среднее время достижения цели или количество полученных наград).
- Подготовьте презентацию, в которой обсудите, как доменные специфики влияют на выбор алгоритмов и стратегий обучения, а также на результаты.

Примерное описание к соревнованию

В соревновании необходимо решить задачу ранжирования по шансу наличия явного фидбека (лайк/дизлайк). Для решения можно использовать любые модели (как затрагиваемые в курсе, так и не встречавшиеся).

В качестве метрики используется средний по всем тестовым пользователям ROC AUC.

Баллы за соревнование (приватный лидерборд):

ROC AUC \geq

0.54 – 1

0.56 – 2

0.58 – 3

0.60 – 4

0.61 – 5

0.62 – 6

0.63 – 7

0.64 – 8

0.65 – 9

0.66 – 10

Примерные задания по тестам

Тест 1.

Вопрос 1.

Где наибольший процент рекомендательного трафика?

- a) Google Play
- b) Amazon
- c) Netflix
- d) Pinterest

Ответ: с.

Вопрос 2.

В 1994 г. GroupLens предложили систему на базе:

- a) Контентного i2i
- b) Коллаборативного i2i

- c) Контентного u2u
- d) Коллаборативного u2u

Ответ: d.

Вопрос 3.

Почему решение участников The Netflix Prize не внедрили?

- a) Большие ансамбли моделей
- b) Смена парадигмы потребления
- c) Смена управления компании
- d) Модели успели устареть

Ответ: a, b.

Вопрос 4.

Рекомендательные алгоритмы оценивают через:

- a) Пользовательские опросы
- b) Онлайн эксперименты
- c) Оффлайн моделирование
- d) Тендеры

Ответ: a,b,c.

Вопрос 5.

Retention пользователей проще всего замерить в:

- a) Длинном онлайн эксперименте
- b) Коротком онлайн эксперименте
- c) Коротком оффлайн моделировании
- d) Длинном оффлайн моделировании

Ответ: a.

Тест 2.

Вопрос 1.

К моделям на графах относятся:

- a) PinSage
- b) SASRec
- c) LightGCN
- d) TwHIN

Ответ: a, c, d.

Вопрос 2.

Implicit ALS отличается от Explicit:

- a) Неявным выводом решения
- b) «Регуляризацией» на отсутствие взаимодействия
- c) Наличием лайков в таргете
- d) Наличием кликов в таргете

Ответ: b.

Вопрос 3.

SASRec отличается от BERT4Rec:

- a) Отсутствием задачи MLM
- b) Способностью работать с последовательностями
- c) Каузальной матрицей внимания
- d) Сдвигом целевой последовательности на 1 позицию

Ответ: a, c, d.

Вопрос 4.

На этапе ранжирования лучше всего работают:

- a) Explicit модели
- b) Implicit модели
- c) Эвристики (свежее, подписки)

d) SLIM

Ответ: а.

Вопрос 5.

К моделям нейросетевого ранжирования относятся:

a) SA-TF

b) HiFormer

c) DCN-V2

d) Wide&Deep

Ответ: b, c, d.

Примерное описание и критерии оценивания к зачету с оценкой

Описание: Экзамен проводится в формате мок-интервью — имитации реального технического собеседования по ML System Design с фокусом на рекомендательные системы. Цель экзамена — проверить умение студента проектировать, анализировать и обосновывать архитектурные решения в области рекомендательных систем, а также демонстрировать глубокое понимание ключевых концепций машинного обучения и инженерных аспектов.

В ходе экзамена студенту будет предложено спроектировать рекомендательную систему для заданного сценария, ответить на вопросы интервьюера, связанные с выбором алгоритмов, обработкой данных, масштабируемостью, объяснимостью, борьбой с смещениями и практическими ограничениями.

Критерии оценивания:

Критерий	Описание	Максимальный балл
Понимание задачи и требований	Способность правильно интерпретировать бизнес-цели и требования системы.	2.0
Архитектура системы	Логичность и полнота предложенной архитектуры, учёт масштабируемости и отказоустойчивости.	2.0
Выбор и обоснование алгоритмов	Обоснование выбора методов рекомендательных систем (матричная факторизация, нейросети и т.д.)	2.0
Обработка данных и качество	Рассмотрение аспектов подготовки данных, борьбы с пропусками и смещениями.	1.5
Объяснимость и интерпретируемость	Умение интегрировать методы объяснимости в систему для повышения доверия пользователей.	1.0
Практические вопросы и ограничения	Учет ограничений вычислительных ресурсов, времени отклика, безопасности и приватности данных.	1.0
Коммуникация и аргументация	Четкость изложения, структурированность ответов и умение вести диалог.	0.5

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Какой способ разделения датасета на трейн/тест в рекомендательных системах наиболее верный? 1. По айтемам 2. По времени 3. По рейтингам 4. По юзерам	2	УК-6
2.	Как называется метрика качества рекомендательной системы, определяющая долю рекомендательных списков, для которых среди верхних K объектов присутствует хотя бы один релевантный элемент?	HitRate / HitRate@K / HitRate at K	ОПК-2
3.	Как называется тип откликов в рекомендательных системах, при котором пользователь напрямую выражает своё отношение к объекту (например, ставит рейтинг или лайк)?	Explicit / Явное / Явное взаимодействие / явное / явное взаимодействие / explicit	ПК-3
4.	Как называется метод матричной факторизации для коллаборативной фильтрации, при котором векторные представления пользователей и объектов оптимизируются поочередно?	ALS / Alternating Least Squares / iALS / implicit Alternating Least Squares	ПК-4