

**Федеральное государственное образовательное
бюджетное учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент анализа данных и машинного обучения
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

Проректор по учебной

и методической работе

_____Е.А. Каменева

25.04.2023 г.

Мейханаджян Л. А.

Рекомендательные системы и коллаборативная фильтрация

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
09.03.03 - Прикладная информатика,
ОП «Инженерия данных», ОП «Прикладная информатика»,
ОП «Прикладные информационные системы в экономике и финансах»

*Рекомендовано Ученым советом
Факультета информационных технологий и анализа больших данных
(протокол №31 от 18.04.2023г.)*

*Одобрено Советом учебно-научного
Департамента анализа данных и машинного обучения
(протокол №2 от 29.03.2023г.)*

Москва 2023

Содержание

1. Наименование дисциплины	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.	2
3. Место дисциплины в структуре образовательных программ	4
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	5
5.1. Содержание дисциплины	5
5.2. Учебно-тематический план	6
5.3. Содержание семинаров, практических занятий.....	7
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	8
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	8
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	9
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	11
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	19
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	19
10. Методические указания для обучающихся по освоению дисциплины	20
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	22
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	23

1. Наименование дисциплины

«Рекомендательные системы и коллаборативная фильтрация».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

ОП «Прикладная информатика»

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКП-3	Способность применять методы разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем	<p>1. Демонстрирует знание назначения и функционала типовых модулей корпоративных информационных систем, основные методы разработки приложений на их платформе.</p> <p>2. Владеет методологией разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем.</p>	<p>Знать понятие модуля и его отличие от понятия функции и класса, подход material design к разработке интерфейса приложения.</p> <p>Уметь реализовывать интерфейс программного средства согласно подходу material design, выявлять ограничения на входные данные в зависимости от решаемой задачи и реализовывать проверку входных данных согласно полученным критериям.</p> <p>Знать общепринятые стандарты описания архитектуры программной системы. Программные механизмы, позволяющие модулю ожидать окончания обработки данных другим модулем или программой.</p> <p>Уметь выбирать архитектуру написания программного средства, наиболее оптимально подходящую под решение поставленной задачи, включая подключение внешних модулей. Реализовывать программные модули согласно выбранной архитектуре с учетом возможности подключения внешних модулей.</p>

**ОП «Инженерия данных»,
ОП «Прикладные информационные системы в экономике и финансах»**

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-4	Способность проектировать и создавать интеллектуальные информационные системы, выбирать метод обучения в соответствии с анализом задачи	<p>1. Демонстрирует знание основных понятий машинного обучения и интеллектуального анализа данных, понимание области и границ применимости, основные виды задач.</p> <p>2. Демонстрирует знание популярных инструментальных средств машинного обучения, собирает датасет, строит модели, проводит их анализ и диагностику, делает содержательные выводы.</p> <p>3. Презентабельно демонстрирует результаты анализа данных и машинного обучения в форме, доступной непрофессионалу, структурирует отчет по проведенному анализу.</p>	<p>Знать технологии чтения и обработки больших данных, принципы создания и алгоритмы анализа интеллектуальных информационных систем.</p> <p>Уметь использовать технологии больших данных для создания интеллектуальных информационных систем и соответствующих решений</p> <p>Знать инструментальные средства Python/R. Методики сбора, фильтрации и предсказания данных, принципы построения, обучения и валидации моделей рекомендательных систем.</p> <p>Уметь реализовывать усвоенные методы и алгоритмы для решения практических задач при помощи инструментальных средств Python/R, фильтровать и преобразовывать исходные данные для оптимизации работы программных средств, выбирать алгоритмы решения задачи в зависимости от исходных данных, строить и обучать модели рекомендательных систем, в том числе, с использованием технологий машинного обучения.</p> <p>Знать различные архитектурные подходы к написанию приложения, их достоинства, недостатки и условия применения. Форматы представления информации.</p> <p>Уметь описывать интерфейс программной системы с формализованном виде согласно принятым стандартам. Реализовывать интерфейс программного средства, выявлять</p>

			ограничения на входные данные в зависимости от решаемой задачи и реализовывать проверку входных данных согласно полученным критериям. Проводит описательный анализ данных, интерпретировать их и представлять отчет заинтересованным лицам.
--	--	--	--

3. Место дисциплины в структуре образовательных программ

Дисциплина «Рекомендательные системы и коллаборативная фильтрация» относится к Циклу профиля (элективный) по направлению подготовки 09.03.03 -Прикладная информатика, ОП «Инженерия данных», ОП «Прикладная информатика», ОП «Прикладные информационные системы в экономике и финансах».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Вид учебной работы по дисциплине	Направление подготовки: 09.03.03 – Прикладная информатика					
	ОП «Инженерия данных» (о)		ОП «Прикладная информатика» (о/озо)		ОП «Прикладные информационные системы в экономике и финансах» о/озо/з (ИОО)	
	Часы: 3/108		Часы: 3/108		Часы: 3/108	
	Всего	Семестр 6	Всего	Семестр 7/8	Всего	Семестр 6/7/8
Общая трудоёмкость дисциплины	108	108	108	108	108	108
Контактная работа - Аудиторные занятия	34	34	34	34	50/34/12	50/34/12
Лекции	16	16	16	16	16/16/4	16/16/4
Семинары, практические занятия	18	18	18	18	34/18/8	34/18/8
Самостоятельная работа	74	74	74	74	58/74/96	58/74/96
Вид текущего контроля	контрольная работа	контрольная работа	контрольная работа	контрольная работа	контрольная работа	контрольная работа
Вид промежуточной аттестации	зачет	зачет	зачет	зачет	зачет	зачет

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Обзор предметной области.

Понятие и история развития рекомендательных систем. Основная терминология: пользователи (users), товары (items), рейтинги, предпочтения и рекомендации, атрибуты users и items. Критерии классификации рекомендательных систем: сфера применения, цель, источник рекомендации, внешние факторы, уровень персонализации, надежность и степень конфиденциальности, интерфейс выдачи рекомендации, рекомендательный алгоритм.

Тема 2. Подходы к построению рекомендательных систем.

Content based, collaborative filtering, гибридные модели, их преимущества и недостатки. Методология создания моделей коллаборативной фильтрации, формализация математической постановки задачи, способы оценки сходства пользователей и объектов, приближение матрицей меньшего ранга. Метрики оценки качества рекомендательных систем. Типы метрик в зависимости от объекта оценки и бизнес-целей. Цели оценки моделей (выбор алгоритма, настройка, сравнение) и ограничения, определение «плохой» и «хорошей» рекомендации. Базовые метрики оценки качества предсказанных рейтингов и рекомендаций. Оценка качества рекомендательных систем (Precision, Recall, F1-мера, HitRate, MAE, RMSE, NDCG).

Тема 3. Программная реализация.

Создание и анализ рекомендательных систем, работающих с явными рейтинговыми данными, используя набор инструментов Python (допустимо в том числе использование Surprise). Построение системы коллаборативной фильтрации на основе алгоритмов «Пользователь-Пользователь» или «Товар-Товар». Аппроксимация матрицы рейтингов

матрицей меньшего ранга и использование ее для гибридной рекомендательной системы. Интерпретация релевантных метрик качества в зависимости от бизнес-целей заказчика системы. Прикладная задача: рекомендация фильмов на примере данных проекта MovieLens.

5.2. Учебно-тематический план

ОП «Инженерия данных» (о), ОП «Прикладная информатика» (о, озо), ОП «Прикладные информационные системы в экономике и финансах» (озо)

№ п/п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Обзор предметной области	28	10	6	4	18	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по лабораторным работам.
2.	Подходы к построению рекомендательных систем.	38	12	6	6	26	
3.	Программная реализация	42	12	4	8	30	
	В целом по дисциплине	108	34	16	18	74	Согласно учебному плану: контрольная работа
	Итого в %		31	47	53	69	

ОП «Прикладные информационные системы в экономике и финансах» (о)

№ п/п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Обзор предметной области	36	18	6	12	18	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по
2.	Подходы к построению рекомендательных систем.	38	18	6	12	20	

3.	Программная реализация	34	14	4	10	20	Лабораторным работам.
	В целом по дисциплине	108	50	16	34	58	Согласно учебному плану: контрольная работа
	Итого в %		46	32	68	54	

Институт Онлайн-образования
ОП «Прикладные информационные системы в экономике и финансах» (3)

№ п/п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Обзор предметной области	20	2	1	1	18	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по Лабораторным работам.
2.	Подходы к построению рекомендательных систем.	42	6	2	4	36	
3.	Программная реализация	46	4	1	3	42	
	В целом по дисциплине	108	12	4	8	96	Согласно учебному плану: контрольная работа
	Итого в %		11	33	67	89	

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Обзор предметной области	<p>Понятие и история развития рекомендательных систем.</p> <p>Критерии классификации рекомендательных систем.</p> <p>Основная терминология: пользователи, товары (users, items), рейтинги, предпочтения и рекомендаций, атрибуты пользователей и товаров.</p> <p><i>Рекомендуемые источники:</i></p> <p>а) основная: 8.1.</p> <p>б) дополнительная: 8.1.</p>	<p>Интерактивная форма</p> <p>Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений</p>

<p>Подходы к построению рекомендательных систем.</p>	<p>Построение рейтингов по популярности, средней оценке. Типы контента. Построение рекомендаций на основе рейтингов, предоставленных пользователями для товаров с известными атрибутами. Построение рекомендаций на основе рейтингов, предоставленных пользователями для товаров с неизвестными атрибутами. Метрики, базовые предикторы, регуляризация.</p> <p>Меры близости: корреляция Пирсона, корреляция Спирмена и др. Поиск ближайших «соседей»-пользователей/товаров на основе матрицы рейтингов (на примере фильмов). Расчет числовых метрик качества построенных рекомендательных систем и интерпретация результатов. Бизнесцели: cross-selling, вовлеченность, покрытие, разнообразие, неожиданность и др.</p> <p>Способы измерения эффекта от рекомендательных систем: offline эксперименты, userstudy, online эксперименты.</p> <p>Метрики: MAE, MSE, RMSE.</p> <p><i>Рекомендуемые источники:</i> а) основная: 8.1. б) дополнительная: 8.1. в) Интернет-ресурсы: 9.9.-9.13.</p>	<p>Интерактивная форма Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений</p>
<p>Программная реализация</p>	<p>Построение рекомендаций с помощью алгоритма «User-user» или «Item-item». Уменьшение размерности матрицы. SVD, Funk SVD, SVD++. Разработка и оценка рекомендательной системы с использованием SVD. Извлечение атрибутов на основе программной реализации TF-IDF анализа. Стратегий объединения подходов.</p> <p><i>Рекомендуемые источники:</i> а) основная: 8.1. б) дополнительная: 8.1. в) Интернет-ресурсы: 9.6-9.13.</p>	<p>Интерактивная форма Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений</p>

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Обзор предметной области	Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем (Amazon.com, Netflix)	– работа с текстом лекции, разбор вопросов по теме занятия; изучение рекомендованных к занятию литературных источников;
Подходы к построению рекомендательных систем.	Стереотипные рекомендательные системы. Способы получения и обработки информации о предпочтениях пользователя. Недостатки неперсонализированных систем. Использование нормализации, снижение «зашумленности» рейтингов. k -ближайших соседей (kNN). Базовые метрики оценки качества предсказанных рейтингов и рекомендаций. Общие принципы обучения и валидации моделей.	– работа с текстом лекции, разбор вопросов по теме занятия; – изучение рекомендованных к занятию литературных источников; – настройка программного обеспечения; – разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы
Программная реализация	Математическая модель FunkSVD. Стохастические алгоритмы для разложения матриц. Архитектура гибридных рекомендательных систем.	– работа с текстом лекции, разбор вопросов по теме занятия; – изучение рекомендованных к занятию литературных источников; – настройка программного обеспечения; – разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы – повторение материала и подготовка к зачету

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные вопросы к контрольной работе

1. Как работает user-to-user и item-to-item рекомендательная система?
2. Используя матричные разложения, разработать рекомендации, учитывающие неявные факторы.

3. Определить метрику для оценки качества рекомендаций, которая лучше всего подходит задаче.
4. Построить рекомендации фильмов пользователям, по которым еще нет никаких данных.
5. По заданной матрице предпочтений вычислить похожесть пользователей одним из методов: корреляция Пирсона, косинусное расстояние, расстояние Жаккара, метод ассоциаций.
6. Классификация методом k ближайших соседей.
7. Методы оценки качества рекомендательных систем.
8. Применимость того или иного подхода в зависимости от постановки задачи.

Примерные задания контрольной работы

1. Реализовать и сравнить три варианта рекомендательной системы фильмов на данных, полученных от преподавателя.
2. Реализовать рекомендательную систему по посещениям. Для пользователя u выдать оценку ресурса g ; выдать ранжированный список рекомендуемых ресурсов.
3. Реализовать рекомендательную систему на основе рейтингов с учетом требуемой точности прогнозов.
4. Самостоятельно составить матрицу предпочтений. На основе принципа k ближайших соседей рекомендации для пользователя u .
5. Составить оценки точности метода фильтрации путем сравнения прогнозируемых рейтингов непосредственно с фактическим рейтингом пользователей.
6. Дана матрица предпочтений пользователей. Приблизить исходную матрицу, матрицей меньшего ранга.

Критерии балльной оценки различных форм текущего контроля успеваемости содержится в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения.

Студент освобождается от выполнения лабораторной работы при наличии сертификата о прохождении темы на онлайн платформе из раздела 9. **"Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины"**.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по данной дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. *«Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».*

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, знаний и умений

ОП «Прикладная информатика»

Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКП-3 Способность применять методы разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем	1. Демонстрирует знание назначения и функционал типовых модулей корпоративных информационных систем, основные методы разработки приложений на их платформе.	<u>Знать</u> Понятие модуля и его отличие от понятия функции и класса, подход material design к разработке интерфейса приложения. <u>Уметь</u> Реализовывать интерфейс программного средства согласно подходу material design, выявлять ограничения на входные данные в зависимости от	Реализовать отдельно интерфейс и модули чтения, обработки данных, обработки ошибок и вывода программы, считывающей из файла, либо с клавиатуры введенную матрицу предпочтений, заполняющий недостающие значения и определяющий меру «похожести» пользователей методом косинусной меры, коэффициента корреляции Пирсона или Евклидовым расстоянием (метод выбирается пользователем после загрузки исходных данных). Каждый метод реализуется в отдельном модуле. Любой модуль программы может быть заменен

	<p>2. Владеет методологией разработки приложений в сфере экономики и финансов на платформе корпоративных информационных систем.</p>	<p>решаемой задачи и реализовывать проверку входных данных согласно полученным критериям.</p> <p><u>Знать</u> Общепринятые стандарты описания архитектуры программной системы. Программные механизмы, позволяющие модулю ожидать окончания обработки данных другим модулем или программой.</p> <p><u>Уметь</u> Выбирать архитектуру написания программного средства, наиболее оптимально подходящую под решение поставленной задачи, включая подключение внешних модулей. Реализовывать программные модули согласно выбранной архитектуре с учетом возможности подключения внешних модулей.</p>	<p>внешней программой с заданным форматом входных и выходных значений.</p> <p>Разработать программный комплекс, считывающий комментарии на сайте магазина, выделяющий наиболее активных пользователей и составляющий их матрицы предпочтений товаров, переводя их комментарии в числовую оценку товара согласно выделенным ключевым словам.</p>
--	---	---	---

**ОП «Инженерия данных»,
ОП «Прикладные информационные системы в экономике и финансах»**

Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКН-4 Способность проектировать и создавать интеллектуальные информационные системы, выбирать метод обучения в соответствии с анализом задачи	1. Демонстрирует знание основных понятий машинного обучения и интеллектуального анализа данных, понимание области и границ применимости, основные виды задач. 2. Демонстрирует знание популярных инструментальных средств машинного обучения, собирает датасет, строит модели, проводит их анализ и диагностику, делает содержательные выводы.	<u>Знать</u> Технологии чтения и обработки больших данных, принципы создания и алгоритмы анализа интеллектуальных информационных систем. <u>Уметь</u> Использовать технологии больших данных для создания интеллектуальных информационных систем и соответствующих решений <u>Знать</u> Инструментальные средства, Python/R. Методики сбора, фильтрации и предсказания данных, принципы построения, обучения и валидации моделей рекомендательных систем. <u>Уметь</u> Реализовывать усвоенные методы и алгоритмы для решения	Группа студентов строит рекомендательные системы на основе предоставленных преподавателем данных (данные могут содержать шумы, выбросы, пропущенные значения и требовать дополнительной обработки, а также формирования синтетических признаков). Результат коллективной работы – выдача рекомендаций, качество которых оценивается по заранее определенному критерию (метрике качества). Исходный полный набор данных разделен на три набора: набор, предназначенный для обучения модели (информация о рейтингах доступна); публичный тестовый набор данных (информация о рейтингах не доступна); закрытый тестовый набор данных (информация о рейтингах не доступна). На основе предоставленных преподавателем данных (данные могут содержать шумы, выбросы, пропущенные значения и требовать дополнительной обработки, а также формирования синтетических признаков) действий пользователей в различных ситуациях, заполнить неизвестные действия для пользователей и предсказать действие в следующей ситуации, для которой известна только ее таблица корреляции с ситуациями из таблицы пользователей.

		<p>практических задач при помощи инструментальных средств Python/R, фильтровать и преобразовывать исходные данные для оптимизации работы программных средств, выбирать алгоритмы решения задачи в зависимости от исходных данных, строить и обучать модели рекомендательных систем, в том числе, с использованием технологий машинного обучения.</p>	
	<p>3. Презентабельно демонстрирует результаты анализа данных и машинного обучения в форме, доступной непрофессионалу, структурирует отчет по проведенному анализу.</p>	<p><u>Знать</u> Различные архитектурные подходы к написанию приложения, их достоинства, недостатки и условия применения. Форматы представления информации.</p> <p><u>Уметь</u> Описывать интерфейс программной системы с формализованном виде согласно принятым стандартам. Реализовывать интерфейс программного средства, выявлять ограничения на входные данные в</p>	<p>По словесной постановке задачи составить матрицу предпочтений с известными оценками пользователей, при возможности дополнить эти оценки информацией по клиенту и товару, и заполнить неизвестные значения. Получить численное выражение «похожести» пользователей одним из методов: косинусная мера, коэффициент корреляции Пирсона, Евклидово расстояние.</p> <p>Разработать и протестировать программное средство, получающее на входе матрицу предпочтений (приложение должно воспринимать ручной ввод, либо чтение из файла в заданном формате) и рассчитывающее коэффициент схожести между пользователями при помощи косинуса угла между векторами рейтингов, корреляции Пирсона либо асимметричного коэффициента Жаккара (выбор производится пользователем). Программное средство должно корректно обрабатывать вводимые</p>

		зависимости от решаемой задачи и реализовывать проверку входных данных согласно полученным критериям. Проводит описательный анализ данных, интерпретировать их и представлять отчет заинтересованным лицам.	данные и предоставлять конечный результат с возможностью фильтрации по пользователям.
--	--	---	---

Примеры заданий

Задача 1. С помощью метода ассоциаций найти похожие фильмы по следующим исходным данным.

	m_1	m_2	m_3	m_4
f_1	1	1	1	
f_2		1		1
f_3	1		1	
f_4			1	1

Задача 2. В таблице приведен набор данных о рейтингах пользователей. Пустые ячейки относятся к элементам, которые пользователи еще не оценили. Используя корреляцию Пирсона, рассчитать коэффициент схожести между пользователями.

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	4		4		5
User 2	4	4			
User 3		3	5	4	5
User 4	3		5		
User 5	3			3	4

Задача 3. В таблице приведен набор данных о действиях пользователей. С помощью асимметричного коэффициента Жаккара рассчитать коэффициент схожести между пользователями.

	Item 1	Item 2	Item 3	Item 4
User 1	1	1	1	
User 2	1		1	1
User 3		1	1	
User 4	1	1		1

Задача 4. В таблице приведен набор данных о рейтингах пользователей. Пустые ячейки относятся к элементам, которые пользователи еще не оценили. Используя косинус угла между векторами рейтингов, рассчитать коэффициент схожести между пользователями.

	Item 1	Item 2	Item 3	Item 4
User 1			3	4
User 2		4		5
User 3	4		4	
User 4			3	4

Задача 5. В таблице приведен набор данных о рейтингах пользователей. Пустые ячейки относятся к элементам, которые пользователи еще не оценили. Спрогнозировать незаполненные ячейки.

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	3	5	4		5
User 2		5		4	
User 3		4	3	4	5
User 4	3		4		

Задача 6. В таблице приведен набор данных о рейтингах пользователей. Пустые ячейки относятся к элементам, которые пользователи еще не оценили. Спрогнозировать незаполненные ячейки.

$$P(1) = \begin{pmatrix} 5 & & 9 & 7 & & 10 \\ 4 & & 3 & & & 8 & 9 \\ & 7 & & 4 & & 9 & \\ 3 & & 6 & 5 & 8 & & 9 \\ & 1 & & & 5 & 5 & 5 \\ & 2 & & 3 & & 5 & 6 \\ 0 & & 0 & 10 & 5 & & 5 \end{pmatrix}$$

Примерные вопросы для подготовки к зачету

1. Понятие и история развития рекомендательных систем.
2. Критерии классификации рекомендательных систем.
3. Стереотипные рекомендательные системы.
4. Неперсонализированные рекомендательные системы.
5. Недостатки неперсонализированных и слабо-персонализированных систем.
6. Способы получения и обработки информации о предпочтениях пользователя.
7. Построение контентно-ориентированной рекомендательной системы.
8. Использование нормализации, снижение «зашумленности» рейтингов.
9. Способы получения и обработки информации о предпочтениях пользователя.
10. Базовые метрики оценки качества предсказанных рейтингов и рекомендаций.
11. Анализ ключевых слов для формирования атрибутов товаров.
12. Общие принципы обучения и валидации моделей.
13. Расширения и альтернативы TF-IDF-анализа.
14. TD-IDF анализ и его производные для выделения ключевых слов.
15. Математическая модель рекомендательной системы.
16. Определение «похожести» интересов пользователей, посчитанное корреляцией Пирсона. Преимущества и недостатки.
17. Построение и актуализация профилей пользователей, взвешивание/нормализация рейтингов.
18. Определение «похожести» интересов пользователей, посчитанное косинусным расстоянием векторов предпочтений. Преимущества и недостатки.
19. Преимущества и недостатки контентно-ориентированного подхода.
20. Определение «похожести» интересов пользователей, посчитанное расстоянием Жаккара. Преимущества и недостатки.

21. Понятие коллаборативной фильтрации.
22. Методики сбора данных.
23. Меры оценки сходства пользователей, преимущества и недостатки.
24. Проблема холодного старта.
25. Алгоритм «User-User» для построения рекомендации на основе оценок пользователей со схожими рейтингами.
26. Актуальность рекомендаций.
27. Использование нормализации, снижение «зашумленности» рейтингов.
28. Подход Slope One.
29. Алгоритм «Item-Item»: рекомендации схожих товаров на основе связей между их рейтингами.
30. Два основных подхода в коллаборативной фильтрации.
31. Ограничения алгоритма «User-User».
32. Переобучение и регуляризация.
33. Двухшаговая реализация алгоритма «Item-Item»: определение сходства товаров на основе их рейтингов.
34. Построение предсказания рейтинга на основе рейтинга «соседей».
35. Гибридные рекомендательные системы.
36. Типы метрик оценки качества рекомендательных систем.
37. Рекомендательные системы, основанные на знаниях.
38. Метрики оценки точности: MAE, RMSE, MSE.
39. Метрики поддержки принятия решений: ROC AUC, полнота и точность.
40. Поиск нерелевантных рекомендаций («reversals»).
41. Задача снижения размерности.
42. Сингулярное разложение матрицы рейтингов (SVD).
43. Математическая модель FunkSVD.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная:

1. Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2022. — 188 с. - ЭБС Лань. — URL: <https://e.lanbook.com/book/198599> (дата обращения: 18.05.2023). — Текст : электронный.

б) дополнительная:

1. Просто о больших данных / Д. Гурвиц, А. Ньюджент, Ф. Халпер, М. Кауфман. — Москва : Эксмо, 2015. — (Библиотека Сбербанка. Т. 58). — ЭБС Сбербанка. — URL: <https://sberbankvip.alpinadigital.ru/book/8007> (дата обращения: 18.05.2023). - Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Информационно-образовательный портал Финансового университета при Правительстве Российской Федерации <http://portal.ufrf.ru/>
2. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
(<http://library.fa.ru/files/elibfa.pdf>)
3. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
4. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
5. Электронно-библиотечная система Znanium <http://www.znanium.com>
6. Электронно-библиотечная система издательства «ЮРАЙТ» <https://www.biblio-online.ru/>
7. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
8. Сайт департамента анализа данных и машинного обучения. http://fa.ru/dep/data_analysis/

9. Массовый открытый онлайн-курс/специализация «Машинное обучение и анализ данных» // МФТИ и Яндекс. – <https://www.coursera.org/specializations/machine-learning-data-analysis>
10. Массовый открытый онлайн-курс/специализация “Recommender Systems” // University of Minnesota – <https://www.coursera.org/specializations/recommender-systems>.
11. Массовый открытый онлайн-курс/специализация «Рекомендательные системы». Introduction to Recommender Systems: Non-Personalized and Content-Based // University of Minnesota – <https://www.coursera.org/learn/recommender-systems-introduction>.
12. Массовый открытый онлайн-курс/специализация “Machine Learning”// Stanford University – <https://www.coursera.org/learn/machine-learning/home/welcome>
13. Профессиональный ресурс. – <https://stackoverflow.com>

10. Методические указания для обучающихся по освоению дисциплины

Раздел содержит комплекс рекомендаций и разъяснений, позволяющих обучающемуся оптимальным образом организовать процесс самостоятельного изучения как теоретического учебного материала, так и подготовки к практическим и семинарским занятиям, в том числе проводимым с использованием активных и интерактивных технологий обучения.

По всем разделам дисциплины предусмотрена цепочка учебных работ: лекция – семинары – самостоятельная работа студентов. По итогам изучения осуществляется аттестация студента в форме письменного контроля знаний (зачета).

Самостоятельная работа студентов в большей мере проходит внеаудиторно. Организации самостоятельной работы служит календарно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного

изучения. На портале <https://campus.fa.ru/> своевременно выкладываются материалы для самостоятельного разбора.

При подготовке к лекции рекомендовано ознакомиться с тематическим планом, в соответствии с рабочей программой дисциплины, и по рекомендованным пособиям выделить наиболее трудные вопросы. Работа с лекционным материалом предполагает конспектирование основного содержания лекции и разбор материала к семинарским занятиям.

Данная дисциплина адаптирована для студентов с ограниченными возможностями здоровья, а также курс состоит из онлайн-части, предоставленной coursera.org (название курса — Introduction to Recommender Systems: Non-Personalized and Content-Based) и автономная часть, описанная РПД. Студенты должны изучить онлайн-часть самостоятельно, используя материалы, доступные на сайте coursera.org. Аудиторные занятия помогут студентам лучше понять основы рекомендательных систем. Содержание дисциплины не ограничивается тематикой онлайн-части и делает особый акцент на актуальных вопросах прикладных направлений, которые могут быть сложными для самостоятельного изучения.

Для студентов Института Онлайн Образования также записаны видеолекции по основным темам дисциплины.

Методические указания по проведению практических занятий

По структуре практические занятия следует разделить на учебные и контрольные.

● **Учебные практические занятия** структурно состоят из следующих компонент.

- 1) проверка наличия выполненного задания самостоятельной работы;
- 2) выборочная проверка корректности выполнения домашнего задания;
- 3) разбор типичных ошибок, возникших в самостоятельной работе;
- 4) рассмотрение теоретических вопросов, связанных с текущим практическим занятием;
- 5) разбор методов выполнения практических заданий и решения задач;

6) корректировка заданий для самостоятельной работы студентов;

7) интерактивная форма – Практикум по решению задач по тематике занятия в малых группах (2-4 студента) – представляет собой выполнение Лабораторных работ, определенных преподавателем, в группе из небольшого количества студентов. Студенты самостоятельно организуют группы. Работа группы оценивается по проценту выполнения.

● Контрольные практические занятия структурно состоят из следующих компонент.

- проведение аудиторной контрольной работы;
- проверка наличия контрольной работы каждого студента;
- разбор типичных ошибок, возникших при выполнении контрольной работы.

При подготовке к практическому занятию рекомендуется повторить или, если это требуется, изучить соответствующий теоретический материал. В случае затруднений обратиться к преподавателю для назначения консультации. Практические занятия проходят, как правило, в интерактивной форме и преподаватель учитывает активность студентов, направленную на решение предложенных задач, и в поиске ответов на вопросы.

Лабораторные работы следует выполнять регулярно при подготовке к практическим занятиям. Если то или иное задание вызвало затруднение необходимо обратиться к преподавателю на консультации или ближайшем практическом занятии.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

Пакет офисных программ

Антивирус Kaspersky

11.2. Современные профессиональные базы данных и информационные справочные системы:

Информационно-правовая система «Консультант Плюс»;

Информационно-правовая система «Гарант»;

Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

Система комплексного раскрытия информации «СКРИН» -
<http://www.skrin.ru>

11.3. Сертифицированные программные и аппаратные средства защиты информации – не предусмотрено

11.4. Статистический пакет R и интерфейс RStudio или другие системы компьютерной математики (например, Matlab).

11.5. Высокоуровневый язык программирования Python.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Мультимедийный класс для проведения лекционных занятий и компьютерный класс для проведения семинарских занятий.