

**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**

СОГЛАСОВАНО

АО «ЦПР»

(наименование организации)

Генеральный директор

(должность представителя работодателя)

_____ Самохин А.Н.

(подпись)

(ФИО)

18.04.2023 г.

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

01.03.02 - Прикладная математика и информатика,

ОП Прикладная математика и информатика,

ОП Прикладное машинное обучение

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.....	12
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	20
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	20
10. Методические указания для обучающихся по освоению дисциплины	21
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	24
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	24

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 «Прикладное машинное обучение»

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКП-2	Способность выполнять анализ качества данных, выявлять и корректировать отклонения в данных и выполнять визуализацию данных	1. Демонстрирует знание методов анализа качества данных для различных типов значений. 2. Владеет методами нормализации данных, в том числе работы с пропусками и выбросами. 3. Владеет современным инструментарием и практическими навыками визуализации данных.	Знать классификацию наиболее актуальных коммерческих и исследовательских инструментов качества данных. Уметь выявлять и исправлять проблемы с данными, которые влияют на точность и эффективность приложений анализа данных. Знать базовые методы восстановления пропусков в массивах данных. Методы приведения различных данных в самых разных единицах измерения и диапазонах значений к единому виду, который позволит сравнивать их между собой (или использовать для расчёта схожести объектов). Уметь применять методы предобработки числовых признаков в обучающих наборах данных с целью приведения их к некоторой общей шкале без потери информации о различии диапазонов. Знать методы предварительной обработки, фильтрации и предсказания данных, методы обобщения и частичного анализа данных. Уметь применять изученные алгоритмы для обработки и анализа больших распределённых массивов данных, корректно объединять полученные частичные решения для получения итогового ответа.

ОП «Прикладная математика и информатика»

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКП-3	Способность составлять прогнозы, готовить рекомендации для принятия финансово-экономических решений	1. Владеет методами прогнозирования, демонстрирует знание принципов и особенностей выбора методов прогнозирования в зависимости от поставленной задачи. 2. Готовит рекомендации для принятия финансово-экономических решений на основе системного подхода для решения поставленных задач.	Знать инструментальные средства MS Excel, Python/R (если применимо). Критерии классификации рекомендательных систем, методики сбора, фильтрации и предсказания данных, принципы построения, обучения и валидации моделей рекомендательных систем. Уметь реализовывать усвоенные методы и алгоритмы для решения практических задач при помощи инструментальных средств MS Excel, 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. Знать различные алгоритмы решения задач рекомендательных систем, реализации этих алгоритмов с применением простейший арифметических функций и функций работы с памятью. Уметь оценивать алгоритм с точки зрения времени работы и задействованной памяти, выбирать и реализовывать для решения поставленной задачи наиболее оптимальный по времени работы и требуемой памяти алгоритм.
ПКП-5	Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического	1. Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	Знать основные способы чтения и обработки больших объемов данных в том числе в условиях ограниченного дискового пространства. Технологии чтения и обработки больших данных, принципы создания и алгоритмы анализа интеллектуальных информационных систем.

	обоснования финансово-экономических решений	2. Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	Уметь работать с технологиями больших данных и алгоритмами решения задач интеллектуальных информационных систем. Знать основы машинного обучения и инструментов языка Python. Уметь решать прикладные задачи методами машинного обучения. Планировать и анализировать результаты А/В-тестов.
--	---	---	--

3. Место дисциплины в структуре образовательных программ

Дисциплина «Рекомендательные системы и коллаборативная фильтрация» относится к Циклу профиля (элективный) по направлению подготовки 01.03.02 - Прикладная математика и информатика, ОП «Прикладная математика и информатика».

Дисциплина «Рекомендательные системы и коллаборативная фильтрация» является дисциплиной Профиля прикладное машинное обучение по направлению подготовки 01.03.02 - Прикладная математика и информатика, ОП «Прикладное машинное обучение».

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

ОП «Прикладное машинное обучение»

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоемкость дисциплины	4 з/е, 144 ч.	144
Контактная работа - Аудиторные занятия	50	50
<i>Лекции</i>	<i>16</i>	<i>16</i>
<i>Семинары, практические занятия</i>	<i>34</i>	<i>34</i>
Самостоятельная работа	94	94
Вид текущего контроля	Проектная работа	Проектная работа
Вид промежуточной аттестации	Экзамен	Экзамен

ОП «Прикладная математика и информатика»

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоемкость дисциплины	3 з/е, 108 ч.	108
Контактная работа - Аудиторные занятия	34	34
<i>Лекции</i>	<i>16</i>	<i>16</i>
<i>Семинары, практические занятия</i>	<i>18</i>	<i>18</i>
Самостоятельная работа	74	74
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	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. Введение

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 фильтраций, 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). 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 «Прикладное машинное обучение»

№ 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.	Введение	30	8	4	4	22	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по Лабораторным работам.
2.	Подходы к построению рекомендательных систем.	52	20	6	14	32	
3.	Программная реализация	62	22	6	16	40	
	В целом по дисциплине	144	50	16	34	94	Согласно учебному плану: проектная работа
	Итого в %		35	32	68	65	

ОП «Прикладная математика и информатика»

№ 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.	Введение	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	

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
Введение	Понятие и история развития рекомендательных систем. Критерии классификации рекомендательных систем. Основная терминология: пользователи, товары (users, items), рейтинги, предпочтения и рекомендаций, атрибуты пользователей и товаров. <i>Рекомендуемые источники:</i> а) литература: 8.1., 8.2. б) Интернет-ресурсы: 9.9.-9.13	Интерактивная форма Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Подходы к построению рекомендательных систем	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. 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,	Интерактивная форма Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

	покрытие, разнообразие, неожиданность и др. Способы измерения эффекта от рекомендательных систем: offline эксперименты, user-study, online эксперименты. Метрики: MAE, MSE, RMSE. <i>Рекомендуемые источники:</i> а) литература: 8.1., 8.2. б) Интернет-ресурсы: 9.9.-9.13.	
Программная реализация	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. <i>Рекомендуемые источники:</i> а) литература: 8.1., 8.2. б) Интернет-ресурсы: 9.9.-9.13	Интерактивная форма Выполнение Лабораторных работ по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

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
Введение	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	– 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

	рекомендаций. Общие принципы обучения и валидации моделей. Механизмы рекомендаций с помощью нейронных сетей и ограниченных машин Больцмана. Латентно-семантический анализ (LSA).	конкретными задачами лабораторной работы
Программная реализация	Математическая модель FunkSVD. Стохастические алгоритмы для разложения матриц. Архитектура гибридных рекомендательных систем. Прикладные задачи: анализ посещаемости сайтов и рекомендация контекстной рекламы.	– 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/экзамену

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 r ; 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.

Примерные задания проектной работы

Задание 1.

- Сравнить качество рекомендаций и время выполнения ALS (без базовых предикторов - without biases) и SGD (без базовых предикторов) для различных параметров регуляризации (например, 0,001, 0,01,...) и числа факторов (например, 1, 10, 25, 50,...).
- Исправить код ALS так, чтобы была возможность производить расчеты с базовыми предикторами (biases).
- Сравнить все 4 подхода по мере MSE, времени на обучение (training phase) и тестирование (inference phase). Результаты, помимо шагов выполнения в `ipython`, необходимо представить в сводной таблице.

Источник данных: MovieLens 100k. Может быть выбран альтернативный, но не меньший по числу оценок, либо пользователей и предметов (items).

Задание 2.

- Требуется реализовать вычисление ошибки [MAE]и [RMSE] на тестовых данных [Movie Lens](<http://grouplens.org/datasets/movielens/>).
- В качестве данных обучения можно использовать файлы с расширением base, а тестирование качества провести на файле test: пары файлов u1.base и u1.test, ..., u5.base и u5.test. Каждая пара -- это разбиение 80%/20% данных для всех пользователей и на обучающие и тестовые данные.
- Для каждого метода (user-based и item-based) постройте графики зависимости [MAE]и [RMSE] от числа соседей (диапазон от 1 до 100 с разумным шагом).
- Сравните подходы на основе полученных результатов по аналогии с пунктами 1 и 2.
- Как Вы считаете, какие фильмы чаще рекомендуются - популярные с высокими оценками или редкие (те, которые редко оцениваются) с высокими оценками?
- Что делать, если соседей (то есть похожих на целевого пользователя или конкретный товар) мало? Нужно/можно ли как-то учитывать достоверность таких рекомендаций?
- * Насколько различны списки из top-n рекомендаций. Попробуйте улучшить результаты подбором β для минимизации MAE (RMSE) в гибридных рекомендациях в зависимости от числа соседей.

Критерии бальной 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 «Прикладное машинное обучение»

Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКП-2 Способность выполнять анализ качества данных, выявлять и корректировать отклонения в данных и выполнять визуализацию данных	1. Демонстрирует знание методов анализа качества данных для различных типов значений.	<u>Знать</u> Классификацию наиболее актуальных коммерческих и исследовательских инструментов качества данных. <u>Уметь</u> Выявлять и исправлять проблемы с данными, которые влияют на точность и эффективность приложений анализа данных.	Имеется набор данных с несколькими признаками (причем признаки могут отличаться и по типу распределения, и по диапазону). Исследовать данные с целью выявления проблем и выработки стратегии их решения. Применить различные методы для разрешения обнаруженных проблем (восстановление пропущенных значений, редактирование аномалий, обработка дубликатов и противоречий и т.д.)

	2. Владеет методами нормализации данных, в том числе работы с пропусками и выбросами.	<u>Знать</u> Базовые методы восстановления пропусков в массивах данных. Методы приведения различных данных в самых разных единицах измерения и диапазонах значений к единому виду, который позволит сравнивать их между собой (или использовать для расчёта схожести объектов). <u>Уметь</u> Применять методы предобработки числовых признаков в обучающих наборах данных с целью приведения их к некоторой общей шкале без потери информации о различии диапазонов.	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.
	3. Владеет современным инструментарием и практическими навыками визуализации данных.	<u>Знать</u> Методы предварительной обработки, фильтрации и предсказания данных, методы обобщения и частичного анализа данных. <u>Уметь</u> Применять изученные алгоритмы для обработки и анализа больших распределенных массивов данных, корректно объединять полученные частичные решения для получения итогового ответа.	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.

ОП «Прикладная математика и информатика»

Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКП-3 Способность составлять прогнозы, готовить рекомендации для принятия финансово-экономических решений	1. Владеет методами прогнозирования, демонстрирует знание принципов и особенностей выбора методов прогнозирования в зависимости от поставленной задачи.	<u>Знать</u> Инструментальные средства MS Excel, Python/R (если применимо). Критерии классификации рекомендательных систем, методики сбора, фильтрации и предсказания данных, принципы построения, обучения и валидации моделей рекомендательных систем. <u>Уметь</u> Реализовывать усвоенные методы и алгоритмы для решения практических задач при помощи инструментальных средств MS Excel, 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	Считать данные о предпочтениях пользователей из нескольких файлов (каждый файл содержит строку оценок различных пользователей для фильма, оценки некоторых пользователей могут быть пропущены), имея таблицу корреляции фильмов, определить общее число пользователей, дающих оценки, корректно соотнести оценки с пользователями и заполнить полученные пустоты предсказанными оценками пользователей.

	2.Готовит рекомендации для принятия финансово-экономических решений на основе на основе системного подхода для решения поставленных задач.	технологий машинного обучения. <u>Знать</u> Различные алгоритмы решения задач рекомендательных систем, реализации этих алгоритмов с применением простейший арифметических функций и функций работы с памятью. <u>Уметь</u> Оценивать алгоритм с точки зрения времени работы и задействованной памяти, выбирать и реализовывать для решения поставленной задачи наиболее оптимальный по времени работы и требуемой памяти алгоритм.	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).
ПКП-5 Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений	1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.	<u>Знать</u> Основные способы чтения и обработки больших объемов данных в том числе в условиях ограниченного дискового пространства. Технологии чтения и обработки больших данных, принципы создания и алгоритмы анализа интеллектуальных информационных систем. <u>Уметь</u> работать с технологиями больших данных и алгоритмами	Проанализировать цифровые показатели процессов оказания услуг государственным учреждением. По конечному множеству пользователей U и конечному множеству услуг I составить пары $R = \{(u, i)\} \subset U \times I$, для которых известна вещественная оценка предпочтения. Требуется по обучающему набору $(R, \{r_{ui}, (u, i) \in R\})$ построить регрессор $\hat{r}(u, i)$, предсказывающий оценку предпочтения для произвольной пары $(u, i) \in U \times I$. Предложить структуру информационной рекомендательной системы для автоматизации процесса принятия

	2. Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.	решения задач интеллектуальных информационных систем. <u>Знать</u> Основы машинного обучения и инструментов языка Python. <u>Уметь</u> Решать прикладные задачи методами машинного обучения. Планировать и анализировать результаты A/B-тестов.	решений государственным учреждением при оказании услуг. Построить рекомендации на основе рейтингов, предоставленных пользователями для товаров с известными атрибутами. Построить рекомендации на основе рейтингов, предоставленных пользователями для товаров с неизвестными атрибутами. Извлечение атрибутов на основе программной реализации TF-IDF анализа
--	---	---	---

Примерные задания для подготовки к зачету / экзамену

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		

Примерные вопросы для подготовки к зачету / экзамену

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.

Пример экзаменационного билета

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

Вопрос №1. (10 баллов) Показатели Top-N с учетом ранжирования (MRR, nDCG).

Вопрос №2. (15 баллов) Как работает алгоритм Funk SVD?

Вопрос №3. (15 баллов) На портале campus.fa.ru имеются данные об оценках пользователей. По заданной матрице предпочтений User/Item. Определить схожесть пользователей а) Косинусное расстояние; б) Коэффициент Отиаи.

Вопрос №4. (20 баллов) На портале campus.fa.ru имеются данные об оценках пользователей. По заданной матрице предпочтений User/Item. Предсказать оценки для незаполненных ячеек с помощью алгоритма Slope One.

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:

2. 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 (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](https://campus.fa.ru/). 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 направления подготовки 01.03.02 — Прикладная математика и информатика, ОП «Прикладное машинное обучение», (Прикладное машинное обучение) на странице дисциплины на учебном портале <https://campus.fa.ru/> в разделе Проектная работа содержатся требования к отчету. Даны источники данных, указаны типы задач для решения, приведено релевантное программное обеспечение. Указаны сроки выбора задачи и сдачи отчета.

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.