

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра _____ «Прикладная математика и информатика»
(наименование)

_____ 09.04.03 Прикладная информатика
(код и наименование направления подготовки)

_____ Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Разработка алгоритмов анализа успеваемости школьников на основе данных об
оценках для персонализированной выдачи рекомендаций по обучающим мероприятиям»

Обучающийся _____ Е.И. Богданов _____
(Инициалы Фамилия) (личная подпись)

Научный _____ к.э.н., доцент, Т.А. Раченко
руководитель _____ (ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Содержание

Введение.....	4
1 Анализ литературных источников, посвящённых исследованиям современных рекомендательных систем	8
1.1 Введение в основы рекомендательных систем.....	8
1.2 Подход к построению персонализированных рекомендательных систем	12
1.3 Значение рекомендательных систем в контексте образования	17
2 Описание методов и алгоритмов работы, предлагаемой гибридной рекомендательной системы.....	24
2.1 Описание методов и алгоритмов работы гибридных рекомендательных систем.....	24
2.2 Не персонализированные рекомендации	38
2.3 Контентная фильтрация	41
2.4 Коллаборативная фильтрация	48
3 Проектирование алгоритмов гибридной рекомендательной системы	51
3.1 Стек технологий, использованный при реализации, предлагаемой гибридной рекомендательной системы	51
3.2 Архитектура предлагаемой гибридной рекомендательной системы	52
3.3 Структура и взаимодействие с базой данных	54
3.4 Разработка алгоритма персонализированной выдачи рекомендаций	57
4 Тестирование предлагаемой гибридной рекомендательной системы.....	71
4.1 Проверка точности прогнозирования интересов пользователя в гибридной рекомендательной системе	71
4.2 Тестирование способности интеграции гибридной рекомендательной системы в веб-приложения	75

4.3 Анализ результатов проведенного тестирования предлагаемой гибридной рекомендательной системы	85
4.4 Будущее развитие предлагаемой гибридной рекомендательной системы	88
Заключение	93
Список используемой литературы и используемых источников.....	96

Введение

Рекомендательные системы занимают важное место в современной образовательной и информационной среде. Внедрение в образование цифровых платформ и технологий, поддерживаемых рекомендательными системами, открывает уникальные возможности для персонализации обучения, адаптации к потребностям учащихся и повышения эффективности образовательного процесса. Развитие методов создания персонализированных рекомендательных систем, таких как контентная фильтрация и совместная фильтрация, важно в свете стремления создать оптимальную среду обучения для каждого отдельного ученика. Такие системы также приобретают все большее значение в корпоративных платформах обучения, помогая адаптировать курсы и материалы к потребностям отдельных учеников. Все это подчеркивает актуальность и важность исследований и разработок в области образовательных рекомендательных систем.

Актуальность исследования заключается в возрастающей потребности современного образования в персонализированном подходе к обучению, который учитывает индивидуальные особенности и уровень подготовки каждого школьника.

Проблема исследования. Низкая вовлечённость обучающихся в образовательный процесс и отсутствие связи рекомендаций с потребностями учащихся.

Целью магистерской диссертации является разработка и экспериментальная оценка эффективности гибридной рекомендательной системы для образовательной среды с ограниченным объемом данных, направленной на повышение релевантности и разнообразия рекомендаций учебных мероприятий и курсов.

Объектом исследования стал процесс выдачи персонализированных рекомендаций основываясь на данных об оценках школьников.

Предметом исследования является разработка алгоритма для реализации

рекомендательной системы, прогнозирующей интерес пользователя к образовательным мероприятиям.

Гипотеза исследования. Предполагается, что использование гибридного подхода, объединяющего контентные, коллаборативные и не персонализированные методы, позволит повысить точность и разнообразие рекомендаций в условиях ограниченного количества данных о пользователе и мероприятиях.

Для достижения цели магистерской диссертации были поставлены следующие исследовательские задачи:

- провести анализ существующих подходов к построению рекомендательных систем;
- провести анализ методов и алгоритмов необходимых для построения рекомендательной системы и на их основе выбрать подходящий подход к построению собственной рекомендательной системы;
- описать механизм работы рекомендательной системы, выдающей персонализированные рекомендации образовательных мероприятий на основании данных об оценках;
- практически реализовать рекомендательную систему основываясь на данных описанного механизма работы;
- протестировать качество выдаваемых рекомендаций и возможность интеграции разработанной рекомендательной системы в сторонние приложения.

Методологической основой исследования являются работы отечественных и зарубежных специалистов, посвященные проблемам изучения и моделирования рекомендательных систем.

Методы исследования включают:

1. Анализ литературных источников - изучение научных статей, монографий и современных исследований в области анализа образовательных данных, машинного обучения и рекомендательных систем. Это позволило выявить актуальные методы обработки

данных об успеваемости и лучшие практики персонализации обучения.

2. Моделирование и визуализация системы – разработка структурных и алгоритмических схем для представления процессов обработки данных, архитектуры хранения информации и взаимодействия компонентов системы. Для наглядного отображения аналитических результатов применялись современные методы визуализации, позволяющие эффективно интерпретировать закономерности в данных и представлять их в удобном для пользователей формате.
3. Разработка и тестирование алгоритмов – создание и экспериментальная проверка алгоритмов машинного обучения для прогнозирования успеваемости и формирования персонализированных рекомендаций. Тестирование проводилось на реальных данных оценок с оценкой точности и эффективности предложенных методов.
4. Оценка качества рекомендаций – проверка релевантности и полезности выданных рекомендаций с помощью метрик точности (precision@k, recall@k, F1-score) и метрик ранжирования (MRR, HR).

Научная новизна выполненных исследований заключается в разработке и экспериментальной проверке гибридной рекомендательной системы, сочетающей контентные, коллаборативные и не персонализированные методы для формирования персонализированных образовательных рекомендаций в условиях ограниченного объема пользовательских данных.

Практическая значимость исследования заключается в разработке гибридной рекомендательной системы, которая позволяет персонализировать образовательный процесс на основе оценок школьников. Данная система обеспечивает выявление индивидуальных потребностей учащихся, что позволяет рекомендовать наиболее подходящие обучающие мероприятия и оптимизировать учебный процесс, для максимальной эффективности.

На защиту выносятся следующие положения:

1. Алгоритм анализа успеваемости школьников на основе данных об оценках для персонализированной выдачи рекомендаций по обучающим мероприятиям.
2. Гибридная рекомендательная система для анализа успеваемости школьников на основе данных об оценках для персонализированной выдачи рекомендаций по обучающим мероприятиям.

Результаты апробации: Результаты исследования были апробированы в процессе функционального тестирования разработанных алгоритмов анализа успеваемости на основании данных об оценках для персонализированной выдачи образовательных мероприятий на базе МБУ «Школа№16». Проведено тестирование работоспособности алгоритмов, а также оценка их эффективности на практике. Экспериментальные данные продемонстрировали эффективность выдаваемых рекомендаций.

Структура магистерской диссертации. Работа состоит из введения, четырёх глав, пятнадцати под-глав, заключения, 5-ти таблиц, 21 рисунков и списка используемой литературы. Основной текст изложен на 98 страницах.

1 Анализ литературных источников, посвящённых исследованиям современных рекомендательных систем

1.1 Введение в основы рекомендательных систем

Современные информационные технологии играют важную роль в образовательных системах, позволяя персонализировать обучение и предлагать студентам и школьникам индивидуальные рекомендации и материалы. Рекомендательные системы, основанные на анализе успеваемости школьников согласно данным об оценках, становятся все более популярными инструментами для обеспечения персонализированного обучения.

В мире, где огромное количество информации доступно одним щелчком мыши, бывает трудно сориентироваться. От выбора фильма для просмотра до выбора курса в учебном заведении - мы сталкиваемся с бесконечным, на первый взгляд, потоком данных. В этом море информации рекомендательные системы выступают в роли проводника, который позволяет нам инновационно взаимодействовать с контентом.

Знакомство с рекомендательными системами - это погружение в пространство, где алгоритмы взвешивают предпочтения и интересы и предлагают нам наиболее подходящие варианты. Но за этими сложными формулами скрывается более глубокая цель. Сделать нашу жизнь проще, открывая новые книги, находя захватывающие фильмы или образовательные ресурсы, которые соответствуют нашим уникальным потребностям и интересам.

Давайте рассмотрим, как работают эти системы, как они учитывают наши предпочтения и почему они стали неотъемлемой частью нашего цифрового путешествия. Добро пожаловать в захватывающий мир рекомендательных систем!

Определение рекомендательных систем

Рекомендательные системы представляют собой набор алгоритмов и

программных продуктов, которые анализируют предпочтения пользователей и предсказывают, что может им понравиться в будущем. Они широко применяются в различных областях, включая интернет-магазины, музыкальные сервисы, видео-платформы и социальные сети [4].

В статье «Рекомендательные системы» авторами было выделено определение рекомендательной системы, как набора алгоритмов, сервисов или программных продуктов, позволяющих определить предпочтения конкретного человека на основе различных данных, включая личную информацию о пользователе и его запросы в поисковых системах. Результатом работы рекомендательных систем являются персонализированные рекомендации, которые помогают пользователю находить интересующий его контент [15].

Одним из лучших примеров рекомендательной системы является TikTok, который в последние годы находится в центре внимания. Пользуясь приложением первые несколько дней, пользователь видит случайные видео, но затем приложение начинает показывать только те, которые могут быть потенциально интересны пользователю, основываясь на длительности просмотра видео, оставленных комментариях и использовании кнопок «Нравится», «Поделиться» и «Прочее».

Музыкальные приложения также не обделены рекомендательными системами. Один из лучших примеров - Spotify, который "пришел" в Россию чуть более трех лет назад. Через несколько дней после установки, программа выбирает песни, которые могут быть интересны пользователю.

В контексте образования, рекомендательные системы могут быть использованы для анализа успеваемости школьников на основе данных об оценках. Это позволяет определить индивидуальные потребности каждого учащегося и предложить ему соответствующие обучающие мероприятия и материалы. Анализ успеваемости основан на статистической обработке данных об оценках, которые школьники получают за выполнение различных заданий, тестов и экзаменов.

Основные компоненты рекомендательных систем

При проектировании рекомендательных систем для персонализированной выдачи рекомендаций по обучающим мероприятиям могут использоваться разные подходы и компоненты. Одними из основных компонентов любой рекомендательной системы являются:

- сбор данных о пользователе и его действиях
- хранение данных о пользователе и его действиях
- моделирование и обучение рекомендательной системы
- методы рекомендаций

Для лучшего понимания рассмотрим каждый компонент рекомендательной системы подробнее.

Процесс сбора данных о пользователе и его действиях необходим для формирования истории взаимодействия пользователя, основанной на просмотрах, оценках и покупках товаров, или услуг. Также этот компонент позволяет собирать демографические данные, такие как: пол, возраст и место проживания. Совмещение истории взаимодействия пользователя и демографических данных может позволить сформировать более полный профиль отдельно взятого пользователя, позволяя системе подстраиваться под текущие интересы и выстраивать новые предпочтения с течением времени. На основании этих данных рекомендательные системы могут максимально точно предлагать различные товары, услуги или в нашем случае, обучающие мероприятия [16].

В основе хранения данных о пользователе и его действиях лежит грамотно разработанная база данных с помощью, которой можно эффективно хранить, организовывать и обрабатывать информацию, полученную при сборе данных. Существует три основных вида баз данных: реляционная, нереляционная и сетевая. В настоящее время самой популярной базой данных является реляционная или «табличная». «Табличкой» можно назвать такую модель баз данных, в которой будет сохраняться вся информация о содержании базы в отдельных таблицах [3].

В базу данных можно сохранять профили пользователей и элементов, полученные во время сбора данных, что позволит рекомендательной системе давать более точные рекомендации. Профили пользователей включают в себя не только базовую информацию, но и аналитические данные о предпочтениях, поведенческих особенностях и предыдущих взаимодействиях с платформой.

Что касается третьего компонента рекомендательных систем, то использование передовых алгоритмов машинного обучения – это ключ к построению эффективной рекомендательной системы, которая способна адаптироваться к изменяющимся обстоятельствам и динамично реагировать на пользовательское поведение. При взаимодействии пользователя с платформой, на основе анализа привычек, реакций на различные элементы и предпочтения пользователя создаются новые модели машинного обучения. Эти модели способны оптимизировать функционал платформы, чтобы предсказывать потенциально возможные интересы пользователей.

Процесс обучения не заканчивается после создания модели, а поддерживается регулярными обновлениями. Это предполагает автоматическое обновление модели на основе новых данных, полученных в результате взаимодействия с пользователем. Такой подход позволяет системе динамически адаптироваться к изменениям в поведении пользователей, новым тенденциям и потребностям. Обновление модели предполагает совершенствование алгоритма для обеспечения повышения точности прогнозирования и релевантности. Этот процесс является ключевым элементом в поддержании системы на передовых позициях в обеспечении персонализированного и релевантного пользовательского опыта.

Заключительным компонентом любой рекомендательной системы являются методы рекомендаций. К методам рекомендаций относятся четыре вида персонализированных рекомендательных систем, а именно: «

- фильтрация, основанная на контенте (content-based filtering);
- коллаборативная фильтрация (collaborative filtering);
- фильтрация, основанная на знаниях (knowledge-based filtering);

– гибридные рекомендательные системы (hybrid filtering)» [2].

Таким образом в данной части мы внимательно рассмотрели, что из себя представляет рекомендательная система и ее основные компоненты.

1.2 Подход к построению персонализированных рекомендательных систем

Захарова И. и Осипов А. пишут, что: «Рекомендации могут быть персонализированными и неперсонализированными. В первом случае мы знаем что-то о пользователе и основываемся на его предпочтениях, во втором - используем популярные или сопутствующие товары» [4].

В первую очередь разберем неперсонализированные рекомендации, потому что они имеют самую простую реализацию. Здесь потенциальный интерес пользователя определяется просто по среднему значению рейтинга продукта. Чаще всего сервисы работают так: «Если большому количеству людей понравилось, значит понравится и вам». Поэтому, если вы не авторизованы в системе, то сервис использует принцип работы, описанный выше [1].

Рейтинги продуктов также могут отображаться различными способами. Например, звезды рядом с товаром, количество «лайков», разница между положительными и отрицательными голосами (часто используется на форумах), процент высоких оценок или гистограмма общих оценок. Гистограммы - самый удобный метод, но у них есть недостаток - если продукты нужно отобразить в виде списка, их сложно сравнивать и ранжировать между собой.

Персональные рекомендации, в свою очередь, предлагают максимизировать использование различной информации о пользователе, но в первую очередь о его покупках.

Для лучшего понимания рассмотрим каждый вид персонализированной рекомендательной системы в отдельности.

Фильтрация на основе контента

Рекомендательные системы, основанные на фильтрации контента, являются наиболее простыми. При выборе рекомендаций учитывается сходство между товарами, услугами или контентом. Если в рекомендательной системе интернет-магазина используется контентная фильтрация, то для конкретной модели смартфона будут просто показаны похожие модели. Например, одна и та же модель смартфона может иметь другой цвет или другой объем памяти [10].

Пример работы рекомендательной системы с контентной фильтрацией представлен на рисунке 1.

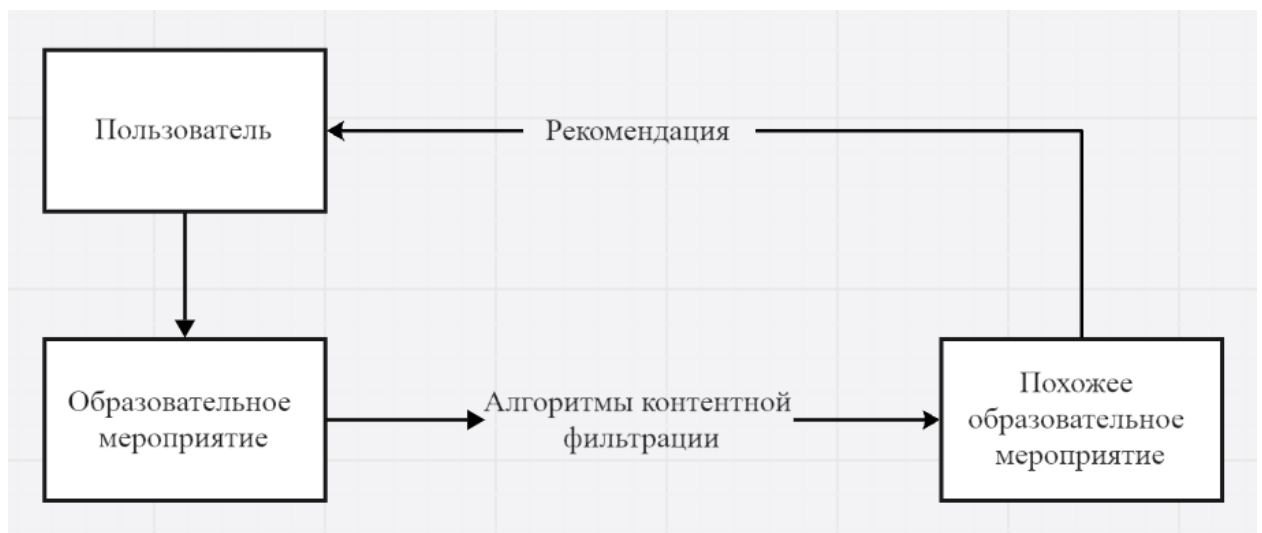


Рисунок 1 – Рекомендательная система с контентной фильтрацией

К преимуществам фильтрации на основе контента относятся предоставления рекомендаций по новым или малоизвестным товарам и независимость от действий других пользователей. Однако она может не справляться с контентом, который трудно объяснить с помощью меняющихся предпочтений и характеристик пользователей.

Коллаборативная фильтрация

Коллаборативная фильтрация является методом рекомендаций, основанным на взаимодействии между пользователями и элементами контента. Он использует отзывы пользователей и данные о предпочтениях,

чтобы предсказать другие элементы, которые могут быть интересны пользователю [11].

В методе коллаборативной фильтрации существует два подхода к обработке информации:

1. Memory-based (на основе памяти). При объединении соседних предпочтений для создания прогнозов используются различные алгоритмы. Предсказания генерируются на основе сходства пользователей (user-based) или сходства предметов (item-based). В первом случае берутся пользователи с похожими предпочтениями и предметы, которые им нравятся, и первому пользователю рекомендуется то, что нравится другим пользователям, и наоборот. Во втором случае сходство вычисляется между предметами, а не между пользователями. Для каждого элемента (товара, курса и т.д.) производится оценка его сходства с другими элементами, основываясь на оценках пользователей. Если пользователя заинтересовал определенный элемент, то система может спрогнозировать, какие потенциально возможные элементы похожие на интересующий элемент могут заинтересовать пользователя в будущем. Сходство между предметами статично, но поскольку предпочтения пользователей могут быстро меняться, подход, основанный на предметах, считается более стабильным и дает лучшие результаты [5, 6].
2. Model-based (на основе модели). Эти методы основаны на машинном обучении и методах интеллектуального анализа данных. Цель заключается в обучении моделей для составления прогнозов. Например, существующие взаимодействия между пользователями и предметами могут быть использованы для обучения модели, предсказывающей пять лучших предметов, которые, скорее всего, понравятся пользователям. Преимущество этих методов заключается в том, что они могут рекомендовать большее количество товаров

большому количеству пользователей по сравнению с другими методами, такими как методы, основанные на памяти. Они утверждают, что имеют большой охват даже при работе с большими разреженными матрицами [23].

Пример работы рекомендательной системы с коллаборативной фильтрацией представлен на рисунке 2.

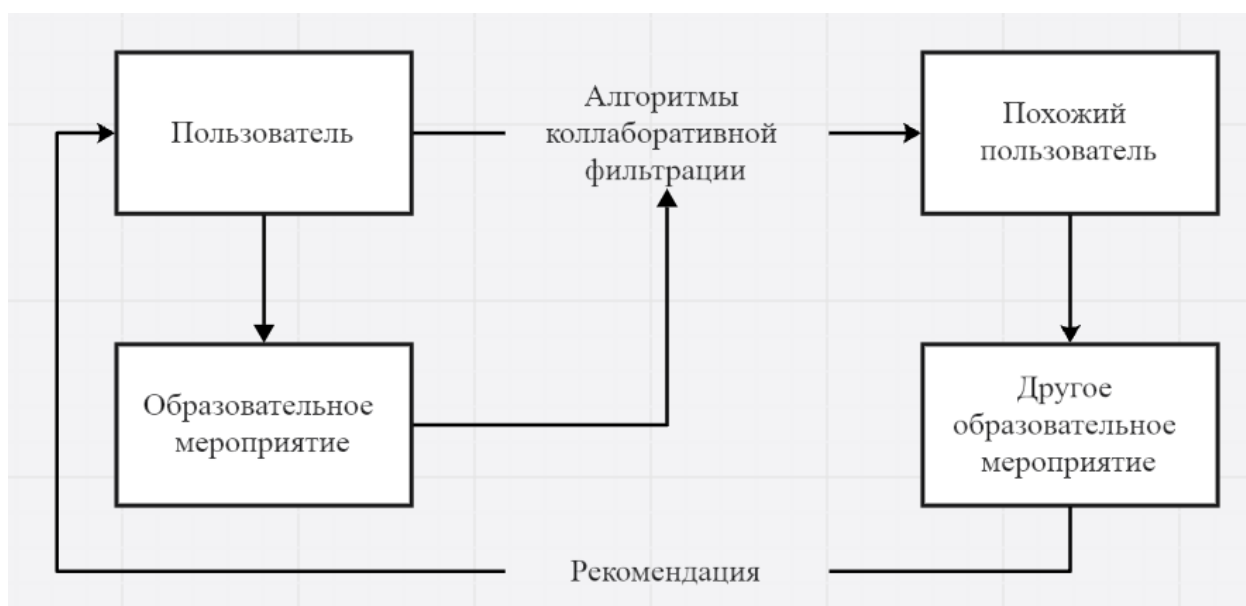


Рисунок 2 – Рекомендательная система с коллаборативной фильтрацией

Преимущества коллаборативной фильтрации в том, что она может предоставлять рекомендации, основанные на реальных предпочтениях пользователей, и не требует predetermined свойств элементов. Однако есть и недостатки, такие как холодный старт для новых пользователей и предметов, проблемы масштабируемости и нехватки данных.

Фильтрация на основе знаний

Системы рекомендаций, основанные на знаниях, особенно полезны, когда товары покупаются нечасто. В качестве примера можно привести жилье, автомобили, финансовые услуги и даже дорогие предметы роскоши. В таких случаях процесс рекомендаций часто страдает из-за отсутствия рейтингов товаров. Системы, основанные на знаниях, не используют рейтинги для составления рекомендаций. Вместо этого процесс рекомендации

осуществляется на основе сходства между требованиями клиента и описанием продукта или с использованием ограничений, определяющих требования пользователя [7].

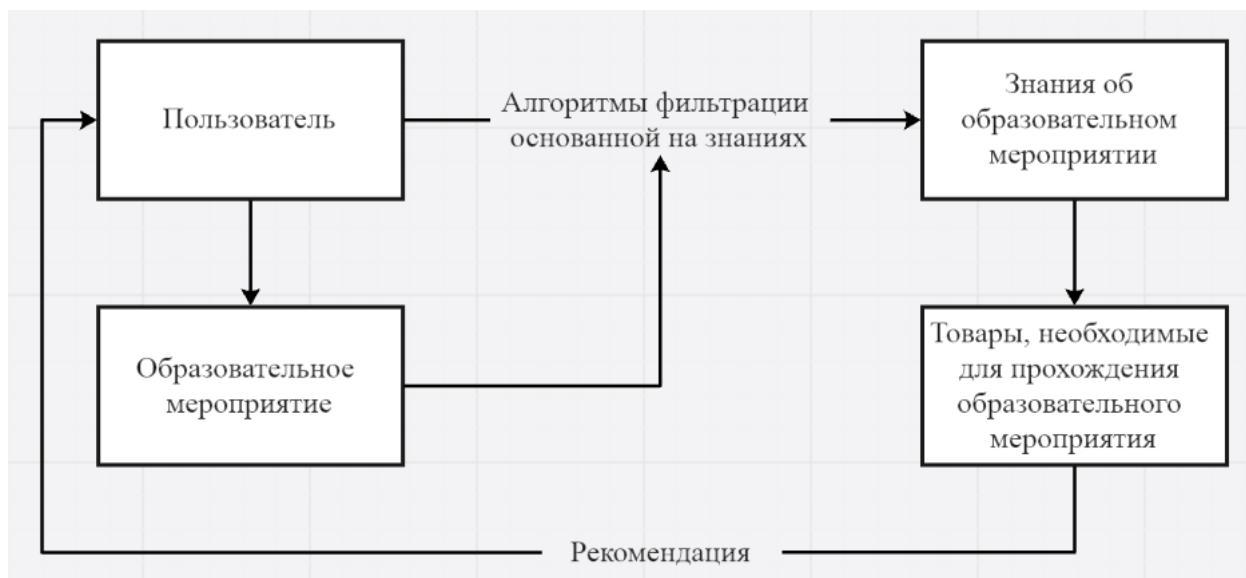


Рисунок 3 – Рекомендательная система с фильтрацией на основе знаний

Характерной особенностью таких систем является то, что пользователи могут четко определить, чего они хотят. Что касается ограничений, то там, где они применимы, они почти всегда применяются экспертами в данной области и известны с самого начала. Например, если пользователь ясно дает понять, что ищет дом в определенном ценовом диапазоне, система должна учитывать этот показатель [10].

Гибридные фильтрации

Гибридные рекомендательные системы еще более сложны, чем предыдущие. Ключевым моментом является то, что два типа фильтрации могут сочетаться одновременно и связываться между собой различными способами:

- раздельная фильтрация и объединение результатов;
- зависит от правил одной фильтрации и другой фильтрации;
- комбинирование двух типов фильтрации.

Вот как это работает на практике. Предположим, что пользователь, похожий на нас, покупает смартфон с синим корпусом, а также чехол и наушники того же цвета. В этом случае в рекомендуемых товарах будет показан синий цвет той же модели смартфона, несколько синих чехлов и синих наушников. Крупные интернет-магазины часто используют гибридную систему.

Таким образом в этом разделе рассматриваются различные подходы к построению персонализированных рекомендательных систем. Контентная фильтрация, коллаборативная фильтрация, фильтрация на основе знаний и гибридные методы предлагают широкий спектр техник для эффективного предсказания и рекомендации. Каждый из этих подходов имеет свои преимущества и недостатки, а выбор конкретной методики зависит от характера данных, целей системы и характеристик пользователя. Эта подглава дает базовое представление о различных стратегиях построения рекомендательных систем.

1.3 Значение рекомендательных систем в контексте образования

Рекомендательные системы выдающие персональные рекомендации по обучающим мероприятиям на основе анализа успеваемости школьников могут сыграть важную роль в образовательном процессе, предоставляя индивидуальные рекомендации и поддерживая учеников на всех этапах обучения. Одна из их ключевых особенностей - способность адаптироваться к индивидуальным потребностям учащихся, учитывая их уровень знаний, интересы и образовательные предпочтения. Это позволяет создать уникальный учебный маршрут, оптимизированный для каждого ученика, что может поспособствовать более эффективному обучению.

Кроме того, рекомендательные системы могут помочь преподавателям оптимизировать процесс обучения. Анализируя данные об обучении, они могут предоставить преподавателям ценную информацию об эффективности

методик, используемых материалов и даже дизайна курса. Это, в свою очередь, позволяет преподавателям оптимизировать свои учебные программы и лучше адаптироваться, и реагировать на меняющиеся потребности учеников.

В образовательном контексте рекомендательные системы могут способствовать формированию сильных учебных сообществ. Предоставляя рекомендации по учебным материалам и заданиям на основе предпочтений учащихся, система может создать более персонализированную среду обучения и повысить мотивацию и вовлеченность в процесс обучения. Таким образом, рекомендательные системы играют важную роль в современной образовательной парадигме, способствуя эффективному и персонализированному обучению.

Применение рекомендательных систем в школьном образовании

К сожалению в настоящее время рекомендательные системы не имеют обширного распространения среди школьного образования, однако инновационные технологии активно внедряются в современный образовательный процесс, и рекомендательные системы не отстают от них. В школьной среде рекомендательные системы предоставляют уникальную возможность персонализировать обучение и оптимизировать учебный процесс.

Одну из таких систем в 2022 году спроектировали специалисты «Яндекс Учебника» [9]. Система предлагала рекомендации заданий по русскому языку и математике для учеников начальной школы. Технология выдавала индивидуальные рекомендации на основе анализа успеваемости учащихся, выявляла проблемы в выполнении заданий, а также их причины, чтобы выдавать более точные рекомендации и повысить эффективность обучения. Система позволяла поэтапно устранять нехватку знаний, исправлять ошибки и выдавать домашнее задание, направленное на улучшение навыков. В персональном аккаунте отображается точная информация об успехах и полученных навыках ученика, что в свою очередь может помочь родителям и учителям направлять процесс обучения и поддерживать детей в развитии и

освоении материала. По словам автора статьи, регулярные занятия в «Яндекс Учебнике» способствуют улучшению знаний учеников на 70-80% от ожидаемого уровня [9].

Как мы видим использование рекомендательных систем и анализа успеваемости школьников на основе данных об оценках имеет ряд преимуществ.

Во-первых, обучение может быть персонализированным и предоставлять индивидуальные инструкции и материалы с учетом уровня знаний, интересов и потребностей учащихся, что способствует более эффективному процессу обучения и повышению успеваемости.

Во-вторых, система рекомендаций может помочь выявить скрытые таланты и потенциал учащихся, предлагая дополнительные учебные мероприятия, соответствующие их интересам и способностям. К таким занятиям относятся внеклассные курсы, олимпиады, научные проекты и другие виды деятельности, способствующие развитию талантов и увлечений учащихся.

В-третьих, анализ успеваемости учащихся на основе данных об успеваемости может помочь выявить проблемы и слабые места в обучении каждого ученика. Это, в свою очередь, поможет учителям и родителям принимать более обоснованные решения о том, какие дополнительные меры и ресурсы необходимы для помощи ученикам в преодолении трудностей обучения. Например, если система определит, что у ученика плохая успеваемость по математике, она может предложить дополнительные индивидуальные занятия с репетитором или учебные ресурсы.

Рекомендательные системы иногда могут выдавать ошибочные и не очень полезные рекомендации. Например, при покупке интересующего нас смартфона в дополнение к нему рекомендуется чехол, но он отличается по размерам или может предназначаться для другой модели смартфона. К сожалению, это может быть не единственной проблемой, поскольку рекомендательные системы не совершенны.

Поэтому, при использовании рекомендательных систем и анализа успеваемости школьников следует учитывать некоторые ограничения и проблемы.

Во-первых, система может иметь доступ только к данным оценки и не может из-за этого выдать полное представление о способностях и интересах ученика. Предоставление доступа к дополнительным данным, таким как результаты тестирований, опросов о предпочтениях и т.д., может повысить точность и надёжность рекомендаций.

Во-вторых, рекомендательные системы могут столкнуться с проблемой "холодного старта", когда у системы недостаточно данных для создания персонализированных рекомендаций для новых учеников, которые еще ни разу не работали в приложении где применяется рекомендательная система. В таких случаях система может использовать агрегированные данные или общие критерии для предоставления начальных рекомендаций, которые затем могут постепенно подстраиваться под предпочтения ученика и его успехи [4].

В-третьих, когда в базу данных рекомендательной системы добавляется новый продукт или услуга, еще не существует пользовательских оценок или отзывов. В результате алгоритм рекомендательной системы не может понять, кому он должен рекомендовать продукт. Такая ситуация известна как "задержка", и в этом случае рекомендательная система будет фокусироваться только на характеристиках продукта до тех пор, пока не накопятся рейтинги и отзывы [14].

В-четвертых, еще одна проблема возникает, когда пользователи не хотят выставлять оценку или писать отзыв о продукте. В терминах науки о данных это называется «разреженностью данных». Системные алгоритмы сталкиваются со сложностью понимания, каким группам пользователей и какие товары нужно предлагать, исходя из схожести их предпочтений. В таких случаях сайты и приложения предлагают вознаграждение за оценки и отзывы, чтобы стимулировать пользователей [10].

Защита конфиденциальности и безопасности данных студентов - это еще один важный момент, потому чтобы предотвратить несанкционированный доступ к информации или ее неправомерное использование, при осуществлении сбора и анализа данных об успеваемости необходимо соблюдать соответствующие правила и нормы.

Рекомендательные системы, анализирующие успеваемость учащихся на основе данных оценок, являются мощным инструментом для индивидуализации образования и повышения успеваемости. Они могут определять индивидуальные потребности учащихся, предлагать подходящие курсы и материалы, выявлять проблемы и находить решения. Однако при использовании таких систем следует учитывать ограничения на доступ к данным и защиту конфиденциальности.

Преимущество персонализированных рекомендаций в обучении

Важную роль в современной системе образования играет персонализация, а рекомендательные системы, которые выдают персонализированные рекомендации по обучению являются очень эффективным инструментом. Адаптация учебных материалов и заданий с учетом потребностей учащихся имеет ряд ключевых преимуществ при использовании такого подхода. При использовании традиционного подхода обучения может происходить недооценка индивидуальных различий в скорости обучения, усвоения материала и понимания, по причине того, что используется ориентировка на средний уровень знаний в классе.

Автор статьи «Инновации в рекомендательных списках: новые форматы и перспективы использования» описывает определение «Персонализация рекомендаций» как: «процесс предоставления каждому пользователю индивидуальных рекомендаций, основанных на его предпочтениях, интересах и поведении. Целью персонализации является улучшение пользовательского опыта и повышение вероятности того, что пользователи найдут интересующие их предметы» [7].

Анализ успеваемости отдельно взятого ученика образовательными организациями будет более эффективным, если будут получены персонализированные рекомендации. Подобные системы помогают определить сильные и слабые стороны обучающихся и на основании этой информации сформировать эффективный индивидуальный план обучения. Такой подход позволяет не только определить проблемные места в обучении, но и подобрать теоретические материалы и практические методики, которые будут лучше всего подходить под конкретного ученика и его взгляды.

Мотивацию для обучения можно создать при помощи эффективного индивидуального обучения. Если ученик почувствует внимание к своим достижениям и потребностям со стороны учителя, то его интерес и вовлеченность к образовательному процессу сильно вырастет. Такой индивидуальный подход к обучению способствует постоянной заинтересованности в развитии навыков и улучшении знаний, а также обеспечивает оптимизацию учебного процесса.

Преимуществ у персонализированных рекомендаций очень много, важнейшей из них является адаптивность. Благодаря постоянному мониторингу ситуации в обучении, система может, учитывая потребности ученика, производить корректировку своих рекомендаций. Такой подход позволяет окружить ученика гибкой учебной средой, в которой он сможет эффективно справляться с трудностями и максимально раскроет свой потенциал в обучении.

Внедрение в российских школах национальных цифровых платформ, таких как "СберКласс", "ЯКласс" и "Яндекс.Учебник", сделало возможным персонализированное и адаптивное обучение. Однако, несмотря на использование технологий в управлении школой, исследования показывают, что применение цифровых методов в обучении ограничено. Персонализация учебного процесса, изменившая модель образования, по-прежнему присутствует в основном в ресурсных и частных школах. Такие учреждения, как гимназия "Универс" в Красноярске и лицей НИУ ВШЭ, являются

примерами успешной персонализации в старших классах и подчеркивают важность культуры взаимодействия для успешной реализации персонализации в образовании [13].

Таким образом в данной под-главе рассматривается значение рекомендательных систем в образовании - области, где рекомендательные системы могут играть важную роль в улучшении образовательного процесса. Применение рекомендательных систем в школьном образовании дает возможность персонализировать обучение и адаптировать учебный процесс к потребностям каждого ученика. Особое внимание было уделено преимуществам персонализированных рекомендаций в обучении, которые не только повышают эффективность образовательного процесса, но и создают мотивацию для учащихся.

Вывод по первой главе.

Рекомендательные системы представляют собой эффективный инструмент персонализации, основанный на предпочтениях и истории взаимодействия пользователей. В главе рассмотрены ключевые компоненты и подходы к созданию таких систем с акцентом на их применение в образовании. Особое внимание уделено школьному обучению, где персонализированные рекомендации способствуют индивидуализации и повышению мотивации учащихся. Представленный обзор закладывает основу для дальнейших исследований в этой области.

2 Описание методов и алгоритмов работы, предлагаемой гибридной рекомендательной системы

2.1 Описание методов и алгоритмов работы гибридных рекомендательных систем

Гибридные система рекомендаций сочетают в себе несколько методов для повышения точности и актуальности рекомендаций. Главной задачей гибридных рекомендательных систем, является преодоление недостатков отдельных методов, таких как коллаборативная фильтрация, контентная фильтрация и не персонализированные рекомендации.

В начале работы с учеником системе не хватает информации о его интересах, для составления списка рекомендаций, такой феномен называют «холодный старт» и для решения проблемы холодного старта может быть использован не персонализированный подход [24].

Авторы статьи «A collaborative filtering recommendation framework utilizing social networks» описывают проблему «холодного старта» следующим образом: «Система ничего не знает о новом пользователе или элементе, когда впервые сталкивается с ним, потому что она не знает историю пользователя или пользовательские оценки элемента. В результате система не может предоставить пользователю рекомендательный сервис и пользователю предоставляется товар, который сложно рекомендовать системе. Для решения этой проблемы характерно использование предложений смешанного использования (например, возраст пользователя, отношения взаимодействия с пользователем, теги продукта и т. д. Идеи смешанного использования объединяют рейтинги и информационное наполнение)» [19].

В нашем случае можно использовать не персонализированные рекомендации для привлечения внимания к курсам с высокой оценкой или посещаемостью.

После определенного количества взаимодействий ученика с понравившимися образовательными мероприятиями, начинает свою работу контентная фильтрация и предлагает образовательные мероприятия, соответствующие уровню знаний и интересам ученика. Под конец коллаборативная фильтрация предложит образовательные мероприятия, которые были высоко оценены другими учениками с аналогичными интересами, что обеспечит лучшую персонализацию.

Основными методами гибридных рекомендательных систем являются: взвешенное объединение, переключение методов, каскадирование, смешивание, мета-уровень, увеличение функциональности и совмещение моделей. Ниже рассмотрим каждый метод по отдельности.

Взвешенное объединение (Weighted). Основная идея метода взвешенного объединения состоит в том, чтобы использовать несколько разных моделей, таких как: коллаборативная фильтрация (CB), контентная фильтрация (CBF) и другие для оценки объектов пользователя, а затем объединения этих оценок с использованием взвешенных сумм. Веса объединения этих оценок с использованием взвешенных сумм. Веса определяют вклад каждого алгоритма в окончательную оценку (формула (1)).

$$\text{evaluation}(q, m) = w_1 \times \text{evaluation}_{U_1}(q, m) + w_2 \times \text{evaluation}_{U_2}(q, m) + \dots + w_n \times \text{evaluation}_{U_n}(q, m) \quad (1)$$

где $\text{evaluation}_{U_1}(q, m)$, $\text{evaluation}_{U_2}(q, m)$, ..., $\text{evaluation}_{U_n}(q, m)$ – оценки m для пользователя q , полученные при использовании различных алгоритмов $U_1, U_2, \dots, U_n$;

$w_1, w_2, \dots, w_n$ – веса, присвоенные каждому алгоритму, в сумме не должны превышать единицы.

В своей книге Франческо Риччи, Лиора Рокача, Брача Шапира и Пола Б. Кантора описывают взвешенное объединение следующим образом: «Учитывая количество рекомендательных баллов, полученных с использованием различных методов, например, $R_{CF}(u, d_i)$, $R_{CB}(u, d_i)$ и т.д.,

общий балл может быть получен с помощью функции агрегирования f (формула 2).

$$R(u, d_i) = f(R_{CF}(u, d_i), R_{CB}(u, d_i), \dots) \quad (2)$$

Система использует линейную комбинацию оценок, полученных в результате совместных и основанных на содержании оценок, для выработки своих рекомендаций и корректирует вес в соответствии с предполагаемыми предпочтениями пользователя. Объединение двух или более методов может быть выполнено с использованием ряда функций с различными свойствами и поведением. Использование нелинейных или более сложных функций позволило бы некоторым рекомендателям более точно настроить процесс ранжирования, создавая менее нерелевантные и более точные прогнозы» [27].

Принцип работы, взвешенной гибридной рекомендательной системы представлен на рисунке 4.

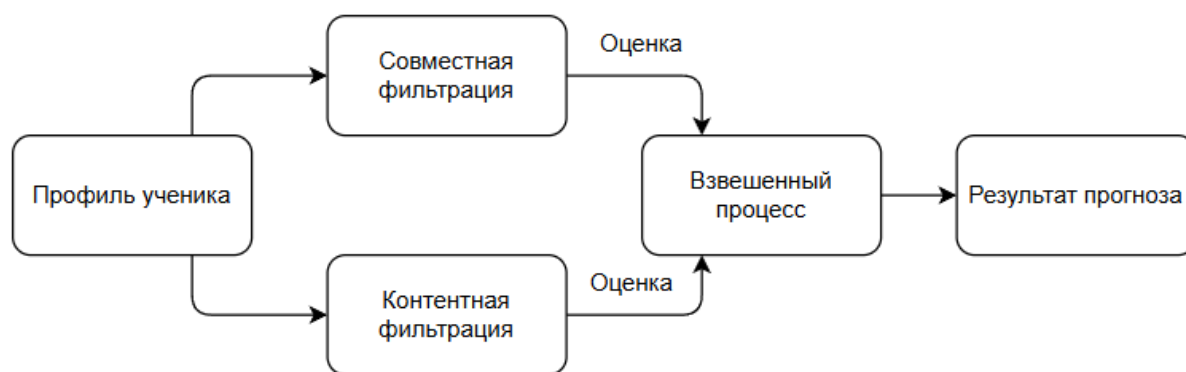


Рисунок 4 - Принцип работы, взвешенной гибридной рекомендательной системы

Переключение методов (Switching). Основная цель метода заключается в улучшении качества рекомендаций за счёт использования наиболее подходящего алгоритма для каждой конкретной ситуации.

В отличие от метода взвешенного объединения, результаты разных алгоритмов напрямую не объединяются. Вместо этого система определяет

алгоритм, который будет использоваться, на основе заранее определенных условий или критериев. Эти критерии могут включать в себя свойства пользователей и объектов, а также доступность данных или производительность алгоритма (формула (3)).

$$V(q, m) = \begin{cases} V_{CF}(q, m), & \text{количество оценок, поставленных учеником} > 6 \\ V_{CB}(q, m), & 1 \leq \text{количество оценок ученика} \leq 6 \\ V_{NRS}(q, m), & \text{количество оценок ученика} = 0 \end{cases} \quad (3)$$

Принцип работы, переключения гибридной рекомендательной системы представлен на рисунке 5.

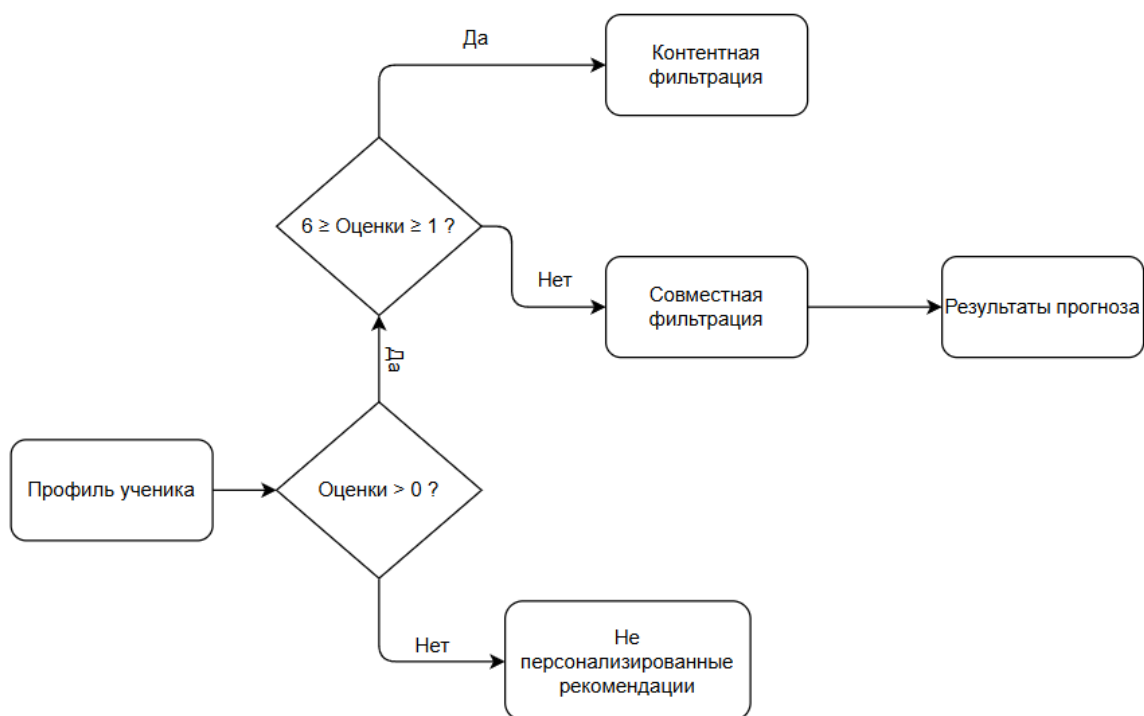


Рисунок 5 - Принцип работы, переключения гибридной рекомендательной системы

Каскадирование (Cascading). Метод каскадирования предполагает использование ряда алгоритмов, каждый из которых дополняет и улучшает

результаты предыдущего алгоритма. На первом этапе используется простой и быстрый метод, который исключает большинство элементов и оставляет только те, которые потенциально актуальны. На более поздних этапах используются более сложные и затрачивающие большое время на вычисление методы для более точного анализа и оценки оставшихся элементов.

В статье «Unit IV Knowledge and Hybrid Recommendation System» автор уточняет: «Рекомендации, полученные с помощью одного метода, используются для улучшения или фильтрации рекомендаций, полученных с помощью другого метода. Например, лучшие рекомендации из системы коллаборативной фильтрации могут быть дополнительно отфильтрованы на основе функций, основанных на контенте» [29].

Допустим, что $V_{NRS}(p)$ является множеством рекомендаций, полученных от не персонализированной рекомендательной системы для пользователя p , $V_{CB}(p)$ является множеством рекомендаций, полученных от рекомендательной системы, основанной на контентной фильтрации для пользователя p , а $V_{CF}(p)$ является множеством рекомендаций, полученных от рекомендательной системы, основанной на коллаборативной фильтрации для пользователя p , тогда финальный набор рекомендаций для пользователя p представлены в формулах (4), (5) и (6):

$$V_{NRS}(p) = NRS(p, i), \quad (4)$$

$$V_{CB}(p) = CB(V_{NRS}(p)), \quad (5)$$

$$V_{CF}(p) = CF(V_{CB}(p)). \quad (6)$$

где p – множества пользователей;

i – множество образовательных мероприятий;

NRS, CB, CF – методы фильтраций.

Принцип работы, каскадной гибридной рекомендательной системы представлен на рисунке 6.

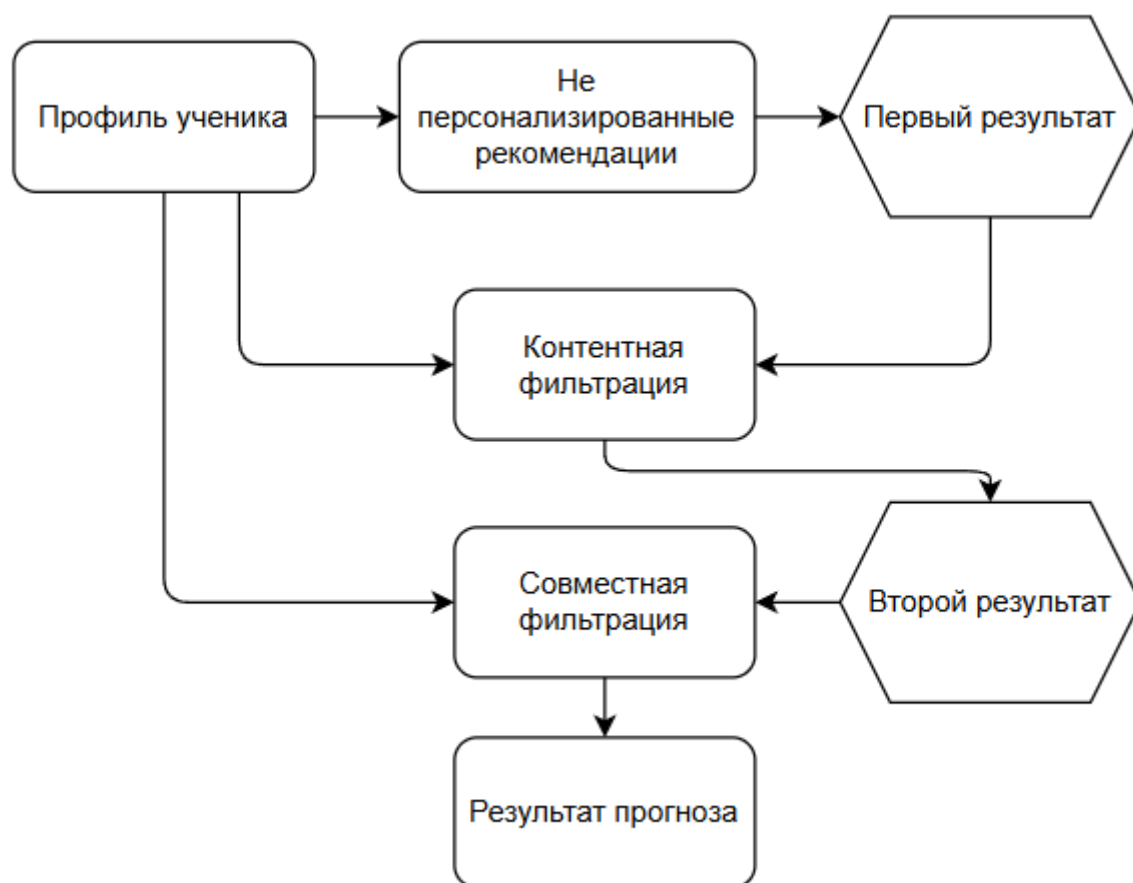


Рисунок 6 - Принцип работы, каскадной гибридной рекомендательной системы

Совмещение характеристик (Feature Combination). Основная цель метода заключается в совмещении различных признаков, полученных в результате использования различных предложенных подходов, в один вектор признаков. Признаки могут включать в себя данные об учениках, обучающих мероприятиях и взаимодействии с ними. Сгенерированный составной вектор признаков используется в обучении модели машинного обучения и составления прогнозов, на основе полученных данных.

При совмещении характеристик каждое взаимодействие предмета и пользователя обозначается, как (p, v) , где p – пользователь, а v – предмет. Все это представляется в виде вектора характеристик $f(p, v)$, который включает в себя сведения о предмете v , пользователе p и взаимодействии $r(p, v)$. При этом,

вектор характеристик может быть представлен следующим образом (формула (7)):

$$f(p, v) = [p, v, r(p, v)] \quad (7)$$

где p – вектор характеристик пользователя u ;

v – вектор характеристик предмета v ;

$r(p, v)$ – данные о взаимодействии предмета и пользователя.

Совмещение характеристик используется для обучения моделей машинного обучения, таких как: линейная регрессия, деревья решений, случайный лес или нейронная сеть. Модель обучается предсказывать целевую переменную (например, оценку или вероятность выбора предмета) на основе вектора характеристик [12].

Модель машинного обучения можно представить в качестве функции f , которая будет принимать на вход вектор характеристик и предсказывать оценку $r(t, d)$ (формула (8)):

$$R(t, d) = f(f(t, d)) \quad (8)$$

Модель обучается на тренировочном наборе данных, используя метод оптимизации (например, градиентный спуск) для минимизации ошибки предсказания. Функция потерь L может быть, например, среднеквадратичной ошибкой между предсказанными и реальными оценками (формула (9)) [30]:

$$L = \frac{1}{K} \sum_{(t,d) \in H} (r(t, d) - \hat{r}(t, d))^2 \quad (9)$$

где K – количество взаимодействий в обучающем наборе данных H .

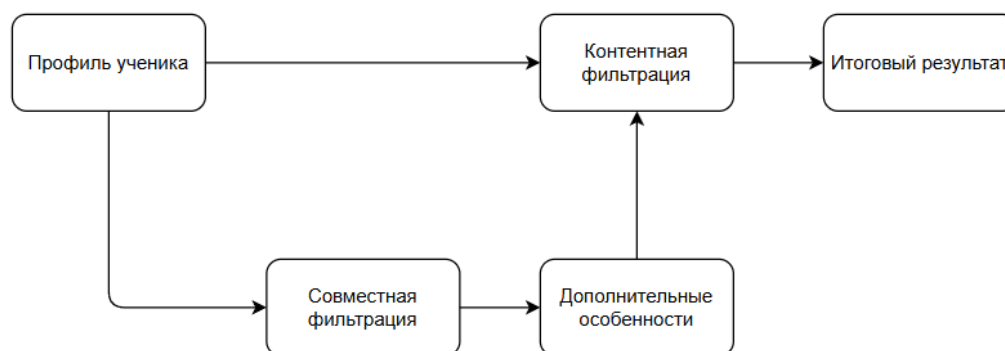


Рисунок 7 - Принцип работы рекомендательной системы, основанной на методе совмещения характеристик

Смешивание (Mixed). Основная цель метода заключается в объединении нескольких разных алгоритмов в один список. В отличие от методов каскадирования и переключения, данный метод использует несколько методов одновременно для генерации персонализированных рекомендаций. Такой подход позволяет учитывать различные аспекты и источники данных, предоставляя более точные и сбалансированные рекомендации.

Пусть у нас есть 3 различных метода рекомендаций: NRS, CB, CF. Каждый метод генерирует свой список рекомендаций $V_{CB}(p)$, $V_{NRS}(p)$, $V_{CF}(p)$ для пользователя p .

Комбинирование списков можно осуществить разными способами. Один из способов состоит в том, чтобы взвесить рекомендации каждого метода. Пусть w_{CB} , w_{NRS} , w_{CF} – веса методов контентной, коллаборативной и не персонализированной фильтрации, тогда формула (10) будет выглядеть следующим образом:

$$V_{mixed}(p) = w_{CB} \times V_{CB}(p) \cup w_{NRS} \times V_{NRS}(p) \cup w_{CF} \times V_{CF}(p) \quad (10)$$

После объединения списков рекомендаций от различных алгоритмов необходимо определить итоговую значимость каждого элемента в результирующем списке. Для этого производится расчёт сводной оценки, отражающей степень соответствия мероприятия интересам конкретного

пользователя. Итоговая оценка вычисляется на основе взвешенных результатов всех использованных методов, как показано в формуле 11:

$$\text{evaluation}(c, p) = w_{CB} \times \text{evaluation}_{CB}(c, p) + w_{NRS} \times \text{evaluation}_{NRS}(c, p) + w_{CF} \times \text{evaluation}_{CF}(c, p) \quad (11)$$

где $\text{evaluation}_{CB}(c, p)$ – оценка предмета c для пользователя p , полученная с помощью метода контентной фильтрации.

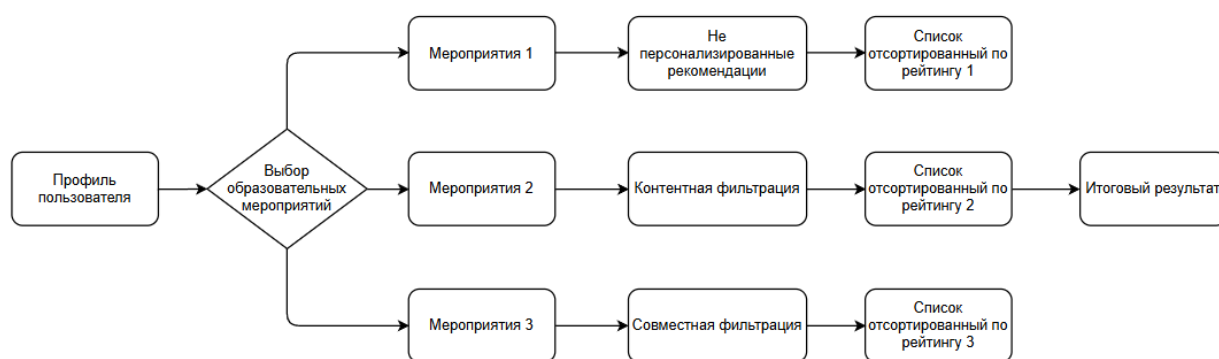


Рисунок 8 - Принцип работы рекомендательной системы, основанной на методе смешивания

Дополнение характеристик (Feature augmentation). В статье под названием «7 типов гибридных рекомендательных систем» о дополнении характеристик автор пишет следующее: «Вспомогательная рекомендательная модель используется для создания рейтинга или классификации профиля пользователя, который в дальнейшем используется в основной рекомендательной системе для получения окончательного прогнозируемого результата.

Гибрид с расширением функций позволяет повысить производительность основной системы без изменения основной рекомендательной модели. Например, используя правило сопоставления, мы можем расширить набор данных профиля пользователя. С помощью расширенного набора данных будет улучшена производительность модели рекомендаций, основанной на содержании» [18].

Данный метод использует те же алгоритмы, что и метод совмещения характеристик (Feature Combination).

Принцип работы рекомендательной системы, основанной на методе дополнения характеристик представлен на рисунке 9.

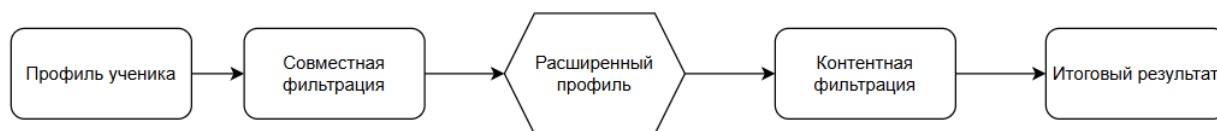


Рисунок 9 - Принцип работы рекомендательной системы, основанной на методе дополнения характеристик

Мета-уровень (Meta-level). В статье под названием «7 типов гибридных рекомендательных систем» о мета-уровне автор пишет следующее: «Гибрид на мета-уровне аналогичен гибриду с дополнением характеристик, в котором вспомогательная модель предоставляет расширенный набор данных для основной рекомендательной модели. В отличие от гибрида с расширением функций, мета-уровень заменяет исходный набор данных изученной моделью из вспомогательной модели в качестве входных данных для основной рекомендательной модели» [18].

Мета-уровневые гибриды часто применяются в системах, где используется глубокое обучение: одна модель обучается для формирования эмбедингов пользователей или объектов, а другая — на этих эмбедингах — формирует финальные рекомендации. Такой способ обеспечивает более гибкую архитектуру, повышает обобщающую способность системы и улучшает качество рекомендаций при наличии сложных и разнородных данных. Вместе с тем, данный подход требует дополнительного этапа обучения и качественного построения промежуточных представлений, от которых во многом зависит эффективность всей рекомендательной системы.

Сравнение всех методов работы гибридных рекомендательных систем представлено в таблице 1.

Таблица 1 - Сравнение методов гибридных комбинаций

Методы	Плюсы	Минусы
Взвешенное объединение	Можно регулировать веса гибридов. Может использоваться в наборах данных с неявной обратной связью.	Ценность конкретных методик рекомендаций должна быть одинаковой для всех алгоритмов. Во всех методиках к каждому элементу применяются весовые коэффициенты, которые могут быть избыточными.
Переключение методов	Система более чувствительна к сильным и слабым сторонам конкретных рекомендательных систем.	Должны быть определены критерии переключения.
Каскадирование	Подходит для тех случаев, когда можно дать большое количество рекомендаций одновременно. Позволяет рекомендовать как популярные, так и новые товары.	Необходимо использовать комбинированную технику. Также должны быть определены правила разрешения конфликтных ситуаций.
Совмещение характеристик	Сочетает в себе функции нескольких алгоритмов, что приводит к улучшению сходства.	Может потребоваться выбор функции в рекомендательной системе, использующей фильтрацию на основе контента.
Смешивание	Позволяет повысить точность системы без ее модификации.	Качество второй рекомендательной системы может зависеть от рекомендаций первой (дополняющей) рекомендательной системы.
Дополнение характеристик	Позволяет использовать вторую рекомендацию только для релевантных элементов (результатов первой рекомендации). Более эффективно, чем взвешенное объединение.	Качество рекомендаций второго уровня может зависеть от рекомендаций первого уровня.
Мета-уровень	Изученная модель - это сжатое представление предпочтений пользователя в отношении элементов.	Качество рекомендаций второго уровня может зависеть от рекомендаций первого уровня.

По итогу проведенного анализа методов, наиболее полезными для создания гибридной рекомендательной системы в нашем случае являются: метод взвешенного объединения, каскадирования и переключения методов.

Далее рассмотрим реализацию каждой части рекомендательной системы по отдельности.

Сбор и предобработка данных об оценках школьников

Сбор и предварительная обработка данных об успеваемости учащихся является важным шагом в создании гибридной системы рекомендаций для анализа успеваемости и предоставления персонализированных рекомендаций. Эти процессы включают сбор данных из различных источников, очистку и преобразование данных для последующего анализа и использования в модели. Давайте подробнее рассмотрим каждый этап.

Источниками данных являются:

- Школьные информационные системы. К ним относятся электронные журналы и дневники с оценками за различные предметы, информация о посещаемости, демографические данные, такие как: пол, возраст, класс;
- Платформы для онлайн-обучения. К ним относятся данные об участии в онлайн-мероприятиях, результаты прохождения тестов;
- Анкетирование и опрос. К ним относятся оценки и отзывы учеников о курсах и учебных материалах, информация о предпочтениях и мотивациях ученика;
- Учительские отчеты и комментарии. К ним относятся качественная оценка успеваемости и поведения учеников, рекомендации учителей по дополнительным занятиям.

Методы сбора данных:

- API интеграция. Данный метод позволяет подключиться к информационной системе образовательного учреждения и образовательной платформе, что позволит автоматизировать сбор данных;

- ETL процессы. Использование ETL (Extract, Transform, Load) инструментов для извлечения данных из разных источников, их преобразование и загрузка в центральное хранилище данных;
- Анкеты и формы. Для сбора дополнительных данных можно создать и провести опрос среди учеников и учителей.

К основным видам предобработки данных относятся: очистка данных, трансформация данных, нормализация данных и обогащение данных. Ниже рассмотрим каждый вид по отдельности.

Во время очистки данных происходит заполнение или удаление записей с отсутствующими значениями, удаление дублированных записей, чтобы избежать искажения анализа, а также исправление ошибок в данных.

Трансформация данных включает в себя преобразование оценок в единую шкалу, суммирование или усреднение оценок по определенным периодам или предметам, а также создание новых признаков, на основе имеющихся данных, таких как средняя оценка по предмету, прогресс за определенный период, участие в дополнительных образовательных мероприятиях.

Нормализация данных включает в себя преобразование значений признаков в определенный диапазон для обеспечения корректной работы алгоритмов машинного обучения и преобразование категорий данных в числовые значения используя методы one-hot и label encoding.

Под обогащением данных подразумеваются процессы интеграции данных из внешних источников, таких как результаты городских экзаменов или данные о среднем уровне успеваемости в регионе.

Функции для анализа успеваемости и выявления образовательных потребностей

После предварительной обработки данных, для того, чтобы вычислить ключевые показатели успеваемости данных системой применяются статистические методы. Ключевые показатели включают в себя такие данные, как: средний балл по каждому предмету, изменения оценок в динамике за

период времени, выявление трендов и распределение оценок среди учащихся. Методы визуализации данных, такие как графики и диаграммы, тоже используются, что позволяет администраторам и преподавателям эффективнее и быстрее ориентироваться в текущей ситуации и выявлять проблемы.

Основываясь на анализе успеваемости учеников, система может выявлять их образовательные потребности. Проводя идентификацию школьных предметов, система определяет, есть ли у учащихся затруднения при изучении предметов или наоборот учащиеся имеют высокий результат. При этом могут использоваться алгоритмы машинного обучения и другие методы анализа данных для определения таких закономерностей и формировании точных персонализированных рекомендаций.

Допустим ученик имеет средний балл 4.5 по химии, в таком случае гибридная рекомендательная система может предложить ему дополнительные материалы по химии или углубленные курсы. Данный процесс можно создать благодаря алгоритмам контентной фильтрации, которые проанализировав содержание доступных образовательных мероприятий, сопоставляют их с успеваемостью и интересами ученика. В то же время учащимся с низкими баллами по определенным предметам могут быть предложены дополнительные уроки, консультации с учителями или интерактивные образовательные программы, которые поспособствуют улучшению их знаний и навыков в конкретных областях.

Одной из ключевых особенностей системы является создание индивидуальных рекомендаций по образовательным мероприятиям. В системе используется гибридный подход, сочетающий в себе контентную и коллаборативную фильтрации. Контентная фильтрация проводит анализ содержания образовательной деятельности ученика и сравниваем ее с его успеваемостью и интересами. Коллаборативная фильтрация использует данные о поведении и предпочтениях других учащихся, на основании чего

создает рекомендации, учитывающие как индивидуальные предпочтения, так и опыт коллектива в целом.

Не менее важным является способность рекомендательной системы к отслеживанию успеваемости учащихся после выданных рекомендаций. Это позволит скорректировать стратегию обучения и адаптировать рекомендации основываясь на новые данные. Благодаря такому подходу можно обеспечить непрерывный учебный процесс, повышая мотивацию учащихся и интерес у обучающихся.

Таким образом в данной части мы внимательно проанализировали все методы и алгоритмы гибридной рекомендательной системы, описали метода сбора и предобработки данных, анализа успеваемости и интересов учеников.

2.2 Не персонализированные рекомендации

Рекомендательные системы, построенные на не персонализированном подходе, преимущественно рекомендуют всем пользователям одно и то же, в следствие чего система продвигает только определенные товары, а пользователь получает не всегда интересующие его рекомендации [25].

Однако для решения проблемы холодного старта не персонализированные рекомендации могут быть очень полезны, так как позволяют гибридной рекомендательной системе получить начальные данные об образовательных мероприятиях для построения будущих персонализированных рекомендаций.

К основным методам не персонализированных рекомендаций относят: популярность контента, средний рейтинг, новинки, выбор редакции, тренды, случайные рекомендации.

Рассмотрим каждый метод не персонализированных рекомендаций по отдельности.

Описание методов и алгоритмов не персонализированной рекомендации

Метод основанный на популярности контента. Популярность контента определяется на основе количественных показателей, отражающих частоту взаимодействия пользователя с определенными элементами системы. Пусть $V(p)$ обозначает популярность элемента i , в таком случае она вычисляется по формуле (12):

$$V(p) = \text{count}(p) \quad (12)$$

где $\text{count}(p)$ – количество взаимодействий с элементом p .

Метод основанный на среднем рейтинге элементов. Рейтинг элемента рассчитывается, как среднее арифметическое оценок, присвоенных учениками. Пусть $R(v)$ будет обозначать средний рейтинг элемента v , в таком случае он вычисляется по формуле (13):

$$R(v) = \frac{\sum_{c=1}^h r_{c,v}}{n} \quad (13)$$

где $r_{c,v}$ – рейтинг, который ученик c присвоил элементу v ;

n – общее количество оценок для элемента v .

Метод основанный на новых элементах. Элементы, недавно добавленные в систему, имеют приоритет над рекомендациями, которые побуждают пользователей пробовать новые образовательные мероприятия. Пусть $T(v)$ обозначает время, когда был добавлен элемент v , в таком случае элементы имеющие наибольшее значение $T(v)$ будут в первую очередь рекомендоваться ученикам.

Метод основанный на выборе редакции. Эксперты или редакторы системы выбирают элементы, которые, по их мнению, имеют наибольшую ценность или актуальность для пользователя. Пусть $E(v)$ обозначает экспертный вес элемента v , в таком случае, элементы имеющие наибольшее значение $E(v)$ будут рекомендоваться ученикам в первую очередь.

Метод основанный на трендах. Пользователям рекомендуются предметы, связанные с текущими событиями и популярными темами, чтобы поддерживать их интересы и информацию в актуальном состоянии. Пусть $TRE(v)$ обозначает актуальные тренды элемента v , в таком случае элементы с наибольшим значением $TRE(v)$ будут рекомендоваться ученикам в первую очередь.

Метод основанный на случайных рекомендациях. Элементы выбираются случайно из доступного списка ресурсов. Пусть $S(v)$ обозначается случайный вес элемента v , в таком случае элемент со случайным значением $S(v)$ будет выдан в качестве рекомендации.

В нашем случае для построения гибридной рекомендательной системы наиболее полезными методами будут: метод основанный на популярности контента, метод основанный на новых элементах и метод основанный на трендах.

Применение не персонализированных рекомендаций в школьном образовании

Введение не персонализированных рекомендаций в школьном образовании значительно улучшит доступность и качество образования в целом для всех учащихся. Не персонализированные рекомендации предоставят одинаковые рекомендации всем пользователям на основе общих показателей, таких как популярность, новинки и тренды. В школьном контексте такие методы могут способствовать распространению высококачественных учебных материалов и поддержанию интереса к обучению.

Например, метод основанный на популярности контента может определять популярные среди учеников образовательные мероприятия основываясь на количестве прохождений или использовании конкретных учебных материалов. А методы, основанные на новинках и трендах, могут автоматически добавлять новые образовательные мероприятия в список рекомендаций по мере их появления.

Таким образом в этом разделе рассматривается реализация не персонализированной фильтрации, анализируются и выбираются подходящие алгоритмы и методы нужные для реализации гибридной рекомендательной системы, а также рассматриваются примеры работы в образовательной организации.

2.3 Контентная фильтрация

Рекомендательные системы, использующие фильтрацию на основе содержания, предполагают, что товары похожие на те, что ранее заинтересовали пользователя также смогут вызвать у него интерес [21]

Содержимое образовательного мероприятия, из которого можно извлечь характерные атрибуты, обычно содержит всю информацию, которая отображается на странице обучающего мероприятия в приложении:

- основная тема и ключевые слова
- описание образовательного мероприятия и учебные цели
- градация обучающих мероприятий по уровню сложности
- тип учебных материалов
- общая длительность курса и объем учебных материалов

Фильтрация на основе контента сравнивает содержание всех образовательных мероприятий с содержанием последнего образовательного мероприятия, который заинтересовал пользователя ранее. В результате такого сравнения появляется коэффициент, отражающий степень сходства образовательных мероприятий, то есть коэффициент интереса потенциального пользователя к образовательному мероприятию.

В своей книге Ким Фальк описывает компоненты, без которых фильтрация на основе контента работать не сможет, а именно: «

- Поведенческий профиль пользователя – ряд таблиц в базе данных, хранящих поведенческую активность пользователя с товарами.

- Анализатор содержания – преобразовывает характеризующую информацию о товаре к удобному для восприятия машины виду, например, к вектору. После чего рассчитывает схожесть между товарами на основе полученного вектора» [22].

На рисунке 7 представлена схема работы фильтрации на основе контента.

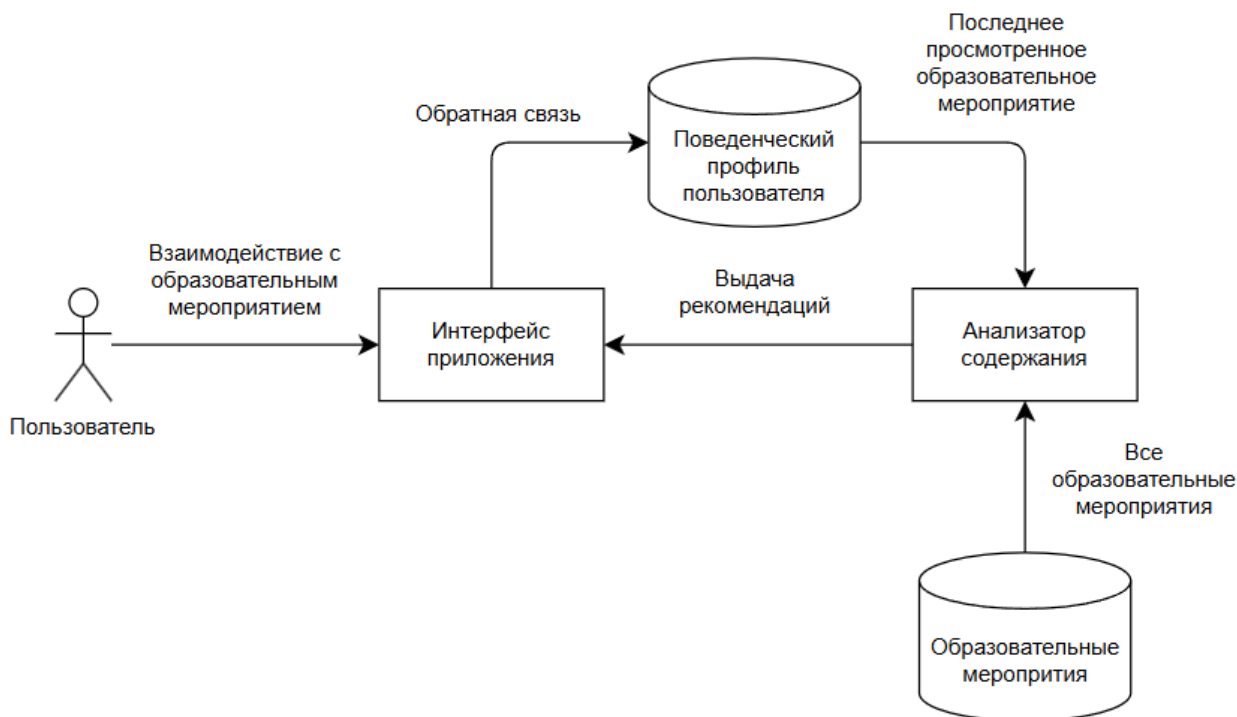


Рисунок 10 - Схема работы фильтрации на основе контента

Основными методами контентной фильтрации являются: извлечение характеристик (feature extraction), векторное представление (vector space model), косинусная близость (cosine similarity), Евклидово расстояние (Euclidean Distance), матричная факторизация (matrix factorization), метод k-ближайших соседей (k-Nearest Neighbors, k-NN), латентное размещение Дирихле (latent Dirichlet allocation, LDA). Ниже рассмотрим каждый метод по отдельности.

Далее рассмотрим реализацию каждой части контентной фильтрации по отдельности.

Поведенческий профиль пользователя

Данный компонент системы рекомендаций использует обучающие мероприятия для отслеживания различных действий пользователя и создания поведенческого профиля.

Профиль поведения пользователя является ключевым компонентом всех систем рекомендаций, основанных на контенте, поскольку система рекомендаций не может генерировать рекомендации для пользователей, не зная, какие образовательные мероприятия ранее интересовали пользователя и как они это делали.

Реализация данного компонента осуществляется путем запоминания действий пользователя с образовательными мероприятиями на уровне кода и хранения этой информации в базе данных, созданной специально для этой цели.

Авторы статьи «Анатомия рекомендательных систем. Часть первая» описывают в своей статье два типа входных данных с которыми приходится работать поведенческому профилю пользователя, а именно: «

- явная обратная связь – представляет собой активности, при которых пользователь в явной форме высказывает свой интерес к товару, например, это может быть оценка или добавление товара в «мне нравится/избранное»;
- неявная обратную связь – определяется путем косвенного высказывания предпочтений пользователя посредством наблюдения за его поведением, например, это может быть просмотр страницы товара или добавление его в корзину» [1].

В предлагаемом решении планируется работать с обоими типами обратной связи, высказанной как явным, так и неявным образом:

- добавление в избранное (явная)
- оценка образовательного мероприятия (явная)
- отзыв образовательному мероприятию (явная)

- просмотр страницы образовательного мероприятия (не явная)
- время, проведенное на странице образовательного мероприятия (не явная)

Явная обратная связь, является более точным отражением интереса пользователя к продукту, но, опять же, не все пользователи снижают рейтинг понравившегося продукта или добавляют его в избранное в небольшом приложении для рекомендации образовательных мероприятий основываясь на оценках учеников. Исходя из этого, имеет смысл также сосредоточиться на неявной обратной связи пользователя. Однако те же рейтинги продуктов могут использоваться для создания рекомендаций не персонализированной системой рекомендаций на основе средних рейтингов продуктов среди пользователей.

Описание методов и алгоритмов контентной фильтрации

Извлечение характеристик (Feature Extraction). Метод не требует специальных формул, но включает использование методов обработки данных и аналитики для создания векторных представлений. Например, TF-IDF часто используется для текстового контента.

Векторное представление (Vector Space Model). Для создания векторного представления текста (например, описания образовательного мероприятия) можно использовать TF-IDF. Такое представление позволяет количественно оценить схожесть между объектами, например, между пользователем и образовательным мероприятием, с помощью метрик расстояния, таких как косинусная мера.

В статье «Системы рекомендаций: обзор современных подходов» Гомзин А.Г и Коршунов А.В. описывают TF-IDF следующим образом:

«Пусть имеется набор текстов (документов). Все тексты разбиты на слова. Коэффициент TF-IDF показывает, насколько слово важно в тексте. Основная идея в том, что слово считается более важным, если оно часто встречается в одном тексте и редко в других.

TF (Term Frequency) - это нормализованная частота слова в тексте (формула (14)):

$$TF(x, D) = \frac{freq(x, D)}{\max_{y \in D} freq(y, D)} \quad (14)$$

Здесь $freq(x, D)$ – количество слов x в документе D . TF принимает значения из отрезка $[0,1]$.

IDF (inverse document frequency) – обратная частота документов формула (15).

$$IDF(x, D) = \frac{freq(x, D)}{\max_{y \in D} freq(y, D)} \quad (15)$$

Здесь N – количество документов в наборе, $n(x)$ – количество документов, в которых встречается слово x . Коэффициент TF-IDF вычисляется как произведение TF на IDF. Для сравнения двух текстов, их можно представить в виде векторов в Евклидовом многомерном пространстве. Каждому измерению соответствует слово, а значение каждой компонент вектора равно TF-IDF. А затем посчитать косинус между этими векторами» [19].

Косинусная близость (Cosine Similarity). Косинусная близость измеряет сходство между двумя векторами.

Так его описывают Гомзин А.Г и Коршунов А.В. в своей статье «Системы рекомендаций: обзор современных подходов»: «Существует множество способов подсчета близости векторов. Один из самых простых – посчитать косинус между этими векторами (формула (16)):

$$\text{sim}(u, a) = \frac{\sum_{i=1}^m r_{a,i} \times r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i}^2} \times \sqrt{\sum_{i=1}^m r_{u,i}^2}} \quad (16)$$

Здесь $\text{sim}(u, a)$ – мера близости (похожести) пользователей a и u . $r_{u,i}$ – значение матрицы R : u строка, i столбец. $\text{sim}(u,a)$ принимает значения из отрезка $[0,1]$.

Если пользователь не указал оценку для какого-то объекта, соответствующее значение матрицы равно 0» [17].

Евклидово расстояние (Euclidean Distance). Евклидово расстояние измеряет расстояние между двумя векторами в многомерном пространстве.

Наиболее распространённая мера расстояния, которую проще всего интерпретировать как длину отрезка, соединяющего две точки. Формула довольно проста, так как расстояние рассчитывается по картезианским координатам точек с использованием теоремы Пифагора [8].

Евклидово расстояние вычисляется по формуле (17):

$$e(v, c) = \sqrt{(v_1 - c_1)^2 + (v_2 - c_2)^2 + \dots + (v_n - c_n)^2} \quad (17)$$

где v и c – векторы характеристик.

Матричная факторизация (Matrix Factorization). В рамках системы рекомендаций декомпозиция матриц включает декомпозицию большой матрицы взаимодействия (например, пользовательских оценок курсов) на 2 меньшие матрицы. Исходная матрица обычно разрежена, потому что не все пользователи оценивают все образовательные материалы. Вы можете использовать матричное разложение, чтобы заполнить пробелы в этой матрице, угадав недостающие значения.

Допустим V является матрицей взаимодействий размером $k \times l$, где k является количеством пользователей, а l количеством объектов. Элемент v_{fc} представляет собой взаимодействие пользователя f с объектом c .

Цель матричной факторизации – разложение матрицы V на две меньшие матрицы:

- S размером $k \times n$, где n – количество латентных факторов для пользователя;
- T размером $l \times n$, где n – количество латентных факторов для объектов.

В конечном итоге получается, что матрица V является произведением двух меньших матриц (формула (18)):

$$R \approx S \times T^G \quad (18)$$

где S и T – матрицы латентных факторов пользователей и объектов.

Метод k -ближайших соседей (k -Nearest Neighbors, k -NN). Основная цель метода заключается в классификации объектов, основанной на близости к другим объектам относительно пространства характеристик. В фильтрации на основании контента данный метод может быть использован для поиска образовательных мероприятий, схожих с теми, которые предпочитает пользователь. Для нахождения ближайших соседей чаще всего используется Евклидово расстояние и косинусная близость между векторами представленные выше в формулах 16 и 17.

Латентное размещение Дирихле (Latent Dirichlet Allocation, LDA). Основная цель метода заключается в извлечении тем из текстового содержания. В контентной фильтрации он используется для определения тем образовательных материалов и сопоставления их с интересами пользователя.

LDA можно представить в виде следующей формулы (19):

$$V(m|n) = \sum_{d=1}^D V(m|c_d)V(c_d|n) \quad (19)$$

где m – слово,

n – документ,

c_d – тема.

По итогу проведенного анализа методов, наиболее полезными для создания гибридной рекомендательной системы в нашем случае являются: метод извлечения характеристик (feature extraction), векторного представления (vector space model), косинусной близости (cosine similarity), Евклидово

расстояния (Euclidean Distance), k-ближайших соседей (k-Nearest Neighbors, k-NN) и латентного размещения Дирихле (latent Dirichlet allocation, LDA).

Таким образом в этом разделе рассматривается реализация контентной фильтрации, анализируются и выбираются подходящие алгоритмы и методы нужные для реализации гибридной рекомендательной системы.

2.4 Коллаборативная фильтрация

Коллаборативная фильтрация анализирует взаимодействия пользователей с учебными материалами для выявления паттернов и предоставления рекомендаций. Этот подход позволяет учитывать схожесть между пользователями или предметами на основе их оценок и поведения [21].

Основными методами коллаборативной фильтрации являются: фильтрация на основе пользователей (user-based) и фильтрация на основе элементов (item-based) и матричная факторизация.

Описание методов и алгоритмов коллаборативной фильтрации

Фильтрация на основе пользователей (user-based). Метод основан на идее, что пользователи, проявляющие схожие предпочтения в прошлом, будут проявлять схожие предпочтения и в будущем. Он определяет пользователей, похожих на текущего пользователя, и рекомендует ему объекты, которые понравились этим пользователям [28].

Фильтрация на основе элементов (item-based). Метод основан на идее, что объекты, которые были оценены одинаково пользователями, имеют схожие характеристики. Он находит объекты, похожие на те, которые оценивал пользователь, и рекомендует их [28].

Матричная факторизация была представлена в под-главе описывающей методы и алгоритмы фильтрации, основанной на контенте.

Фильтрация на основе пользователей и фильтрация на основе элементов используют методы, такие как: косинусная близость, предсказание оценки и корреляция Пирсона.

Косинусная близость. Определение и формула (2.16) представлены в под-главе «2.3.2 Описание методов и алгоритмов фильтрации, основанной на контенте».

Корреляция Пирсона (Pearson Correlation Coefficient). Авторы статьи «Performance Evaluation of Time-based Recommendation System in Collaborative Filtering Technique» описывают метод корреляции Пирсона следующим образом: «Коэффициент корреляции Пирсона (PCC) определяет сходство, принимая во внимание только сопоставимые элементы. Он рассчитывает сходство в диапазоне от -1 до +1. Показатель сходства с помощью PCC рассчитывается следующим образом (формула (20)):

$$\text{sim}(u_a, u_b)^{\text{PCC}} = \frac{\sum_{i=1}^{i'} (r_{u_a,i} - r_{u_a})(r_{u_b,i} - r_{u_b})}{\sqrt{\sum_{i=1}^{i'} (r_{u_a,i} - r_{u_a})^2} \times \sqrt{\sum_{i=1}^{i'} (r_{u_b,i} - r_{u_b})^2}} \quad (20)$$

По сравнению с другими показателями сходства, при которых данные должны быть нормализованы, этот метод хорошо работает, когда данные не нормализованы, что повышает его применимость. Показатель PCC имеет ограничение, заключающееся в том, что он не может найти взаимосвязь между пользователями, если количество общих элементов недостаточно» [26].

Метод предсказания оценки можно представить в виде следующей формулы (21):

$$V_{p,c} = \frac{\sum_{h \in L_k(u)} \text{sim}(p,h) \times v_{h,c}}{\sum_{h \in L_k(u)} |\text{sim}(p,h)|} \quad (21)$$

По итогу проведенного анализа методов, наиболее полезными для создания гибридной рекомендательной системы в нашем случае является фильтрация на основе пользователей (user-based) и его методы, такие как: косинусная близость и корреляция Пирсона.

Таким образом в этом разделе рассматривается реализация коллаборативной фильтрации, анализируются и выбираются подходящие алгоритмы и методы нужные для реализации гибридной рекомендательной системы.

Вывод по второй главе.

Проведя анализ литературы был представлен подробный анализ методов и алгоритмов, используемых в гибридной рекомендательной системе, включая коллаборативную, контентную и не персонализированную фильтрацию. В каждой фильтрации отобраны подходящие алгоритмы, обеспечивающие эффективное представление рекомендаций, основываясь на успеваемости и интересах учеников. Также описаны методы сбора и предварительной обработки данных, обеспечивая прочную основу для системы, которая может сочетать различные подходы для получения точных и полезных образовательных рекомендаций.

3 Проектирование алгоритмов гибридной рекомендательной системы

3.1 стек технологий, использованный при реализации, предлагаемой гибридной рекомендательной системы

При разработке, предложенной гибридной рекомендательной системы использовался язык программирования Python. Данный язык программирования достаточно популярен в области Data Science, а также для него написано большое количество библиотек и фреймворков.

На начальном этапе практической реализации гибридной рекомендательной системы было принято решение ориентироваться на потенциально возможные интеграции с веб-приложениями, в том числе разработанными на основе фреймворка Django.

Основываясь на выбранных решениях был сформирован следующий стек-технологий:

- Django – фреймворк для объединения веб-приложения с рекомендательной системой и базой данных;
- PostgreSQL – база данных в которой хранится информация об образовательных мероприятиях, учениках и их явных и не явных взаимодействиях с рекомендательной системой;
- PgAdmin – для визуализации актуального состояния базы данных;
- Docker – для создания контейнера в котором хранится база данных;
- HTML – язык гипертекстовой разметки для упорядочивания содержания веб-страницы;
- CSS – язык описания внешнего вида веб-страницы;
- Scikit-learn – библиотека для машинного обучения, использовалась для векторизации текста и расчета косинусного сходства;
- Pandas – библиотека для обработки и анализа данных;
- NumPy – библиотека для работы с массивами и матрицами.

– GitHub (для хранения кода)

Программный код писался в интегрированной среде разработки – Visual Studio Code.

Для хостинга исходного кода проекта и хранения информации о всех изменениях в коде был использован сервис – GitHub, основанный на системе контроля версий Git.

С использованием упомянутого набора инструментов, включающего языки программирования, фреймворки и библиотеки, была реализована предложенная гибридная рекомендательная система.

Таким образом в данной части нами был рассмотрен стек технологий, который использовался при создании предлагаемой гибридной рекомендательной системы.

3.2 Архитектура предлагаемой гибридной рекомендательной системы

Архитектура системы описывает структуру и взаимодействие основных компонентов, используемые технологии и принципы проектирования. Она показывает, как модули работают вместе для достижения целей проекта, обеспечивая масштабируемость, надежность и соответствие требованиям.

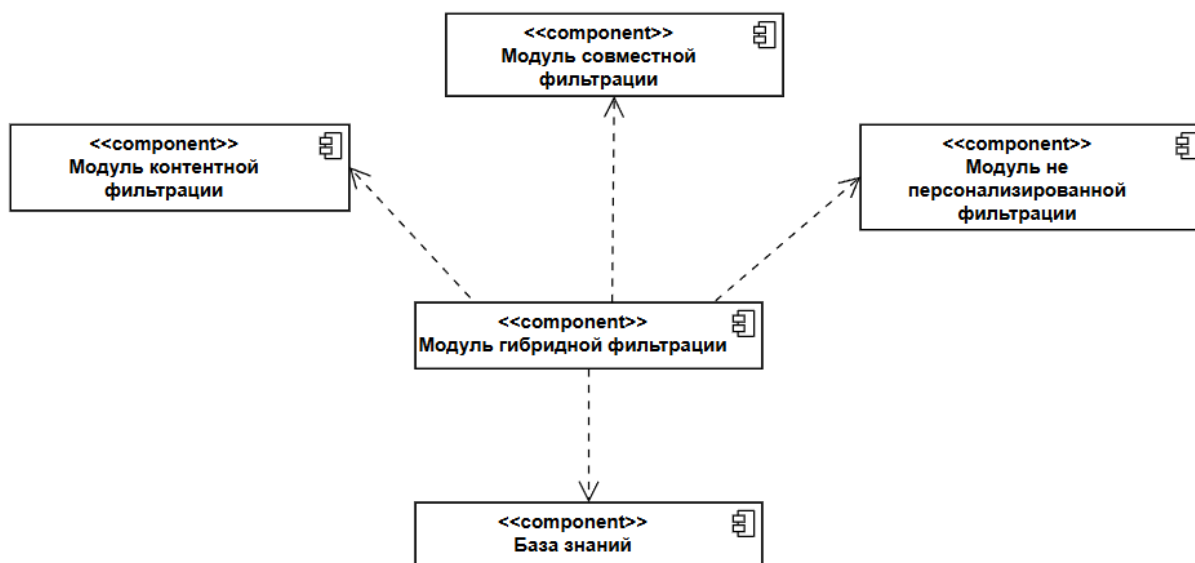


Рисунок 11 - Диаграмма компонентов предлагаемой гибридной рекомендательной системы

Модуль гибридной фильтрации – является центральным компонентом рекомендательной системы в котором происходит генерация рекомендаций путём совмещения разных методов фильтрации, таких как: переключение методов (Switching), каскадирование (Cascading), взвешенное объединение (Weighted).

Модули контентной, совместной и не персонализированной фильтрации используются на разных этапах выдачи рекомендаций и вмещают в себя функционал в соответствии с методом фильтрации.

База данных используется модулем гибридной фильтрации для получения данных об учениках и образовательных мероприятиях, используемых при генерации рекомендаций.

Таким образом в данной части мы наглядно рассмотрели архитектуру предлагаемой гибридной рекомендательной системы.

Далее рассмотрим каждый модуль подробнее.

3.3 Структура и взаимодействие с базой данных

База данных играет ключевую роль в функционировании гибридной рекомендательной системы, обеспечивая хранение и управление данными, необходимыми для генерации персонализированных рекомендаций. В нее включена информация об учащих, их успеваемости, образовательных мероприятиях, а также о явных и неявных взаимодействиях с системой. Структура базы данных представлена на рисунке 12.

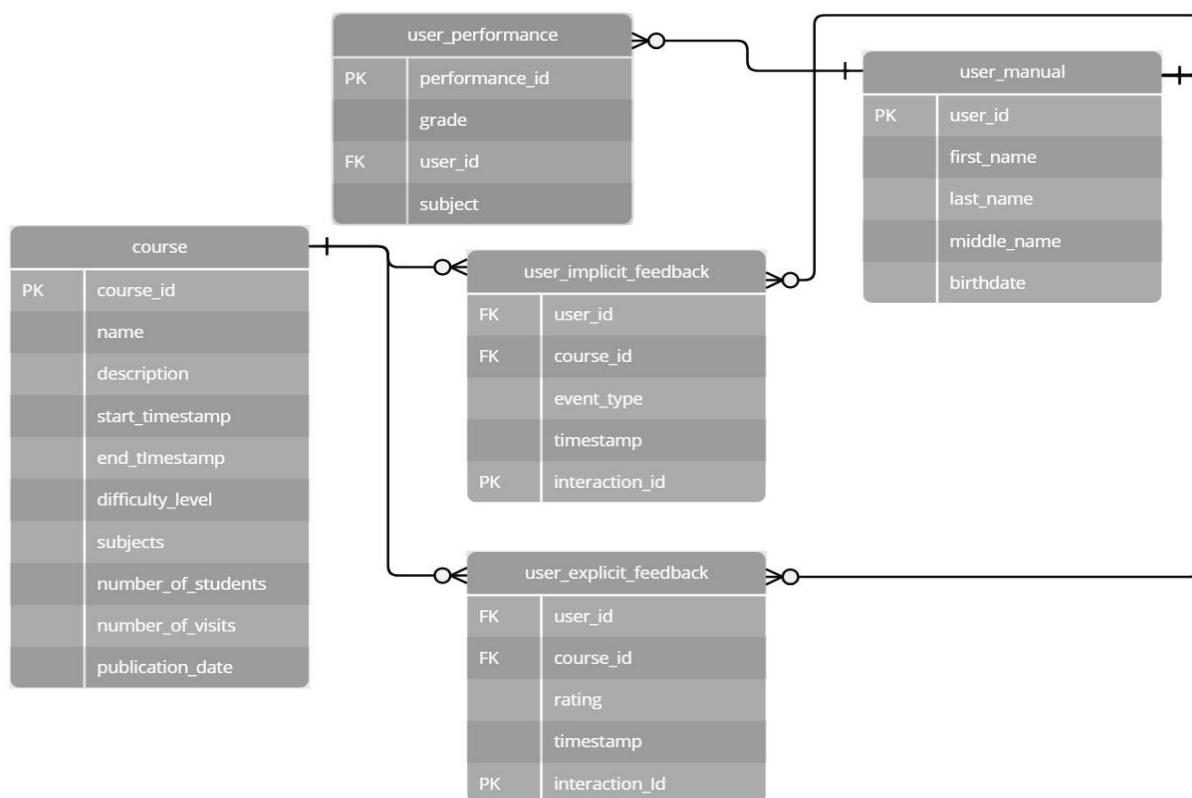


Рисунок 12 – Структура базы данных

База данных спроектирована таким образом, чтобы обеспечить эффективное хранение и обработку информации, необходимой для генерации персонализированных рекомендаций. Она включает в себя несколько ключевых таблиц, каждая из которых отвечает за хранение определенных данных.

Образовательные мероприятия, которые будут рекомендоваться нашей гибридной рекомендательной системой хранятся в таблице «course» и включают следующие данные:

- `course_id` – это поле таблицы, содержащее ID образовательного мероприятия;
- `name` – это поле таблицы, содержащее имя образовательного мероприятия;
- `description` – это поле таблицы, содержащее описание образовательного мероприятия;
- `start_timestamp` – это поле таблицы, содержащее временную метку, которая фиксирует начало проведения образовательного мероприятия;
- `end_timestamp` – это поле таблицы, содержащее временную метку, которая фиксирует конец проведения образовательного мероприятия;
- `difficulty_level` – это поле таблицы, содержащее приблизительную сложность образовательного мероприятия. В нашей системе существуют следующие уровни сложности: `beginner` (начальный уровень), `intermediate` (средний уровень) и `advanced` (высокий уровень);
- `subjects` – это поле таблицы, содержащее предметы, относящиеся к образовательному мероприятию;
- `number_of_students` – это поле таблицы, содержащее максимально допустимое количество участников образовательного мероприятия;
- `number_of_visits` – это поле таблицы, содержащее количество посещений данного образовательного мероприятия;
- `publication_date` – это поле таблицы, содержащее дату публикации образовательного мероприятия на сайте.

Данные о пользователях хранятся в таблице «user_manual» и включают следующие данные:

- user_id – это поле таблицы, содержащее ID пользователей, которые зарегистрированы в система;
- first_name – это поле таблицы, содержащее имя пользователя;
- last_name – это поле таблицы, содержащее фамилию пользователя;
- middle_name – это поле таблицы, содержащее отчество пользователя;
- birthdate – это поле таблицы, содержащее дату рождения пользователя.

Данные об успеваемости пользователей хранятся в таблице «user_performance» и включают следующие данные:

- performance_id – это поле таблицы, содержащее ID оценки успеваемости по предмету определенного пользователя.
- grade – это поле таблицы, содержащее оценку успеваемости по предмету определенного пользователя.
- user_id – это поле таблицы, содержащее ID пользователя, которому принадлежит оценка успеваемости по предмету.
- subject – это поле таблицы, содержащее предмет, по которому была выставлена оценка успеваемости.

Неявные взаимодействия фиксируются в таблице «user_implicit_feedback» и включают следующие данные:

- user_id – это поле таблицы, содержащее ID пользователей, которые взаимодействуют с системой;
- course_id – это поле таблицы, содержащее ID курса, который был целью взаимодействия;
- event_type – это поле таблицы, содержащее типы неявных взаимодействий. В нашей системе существуют следующие типы взаимодействия с курсом: viewed (просмотр курса), enrolled (запись на курс) и completed (закончить курс);

- timestamp – это поле таблицы, содержащее время, когда произошло неявное взаимодействие.

Явные взаимодействия фиксируются в таблице «user_explicit_feedback» и включают следующие данные:

- user_id – это поле таблицы, содержащее ID пользователей, которые взаимодействуют с системой;
- course_id – это поле таблицы, содержащее ID курса, который был целью взаимодействия;
- rating – это поле таблицы, содержащее оценку пользователя о курсе от 1 до 5;
- timestamp – это поле таблицы, содержащее время, когда произошло неявное взаимодействие.

Для передачи данных системе были разработаны отдельные «репозитории» для каждой таблицы в базе данных, которые вмещают в себя операции чтения и сохранения информации. Каждый из «репозиторий» имеет класс-сущность для трансформации в тип данных, понимаемый системой.

Таким образом в данном разделе описывается организация структуры базы данных, принципы взаимодействия между ней и компонентами системы.

3.4 Разработка алгоритма персонализированной выдачи рекомендаций

Модуль не персонализированной фильтрации

Система не может выдавать персонализированные рекомендации, пока нет никаких данных.

Для решения проблемы «холодного» старта в системе используется модуль не персонализированных рекомендаций, который представлен следующими функциями:

- getMostPopularCourses - сортирует курсы по количеству их посещений в убывающем порядке и выдаёт определенное

количество самых популярных курсов, число которых определяется константой.

- `getMostRecentCourses` - сортирует курсы в убывающем порядке по дате публикации.

Модуль не персонализированной фильтрации предоставляет пользователям доступ к самым популярным и актуальным образовательным ресурсам даже без наличия индивидуальных данных. Однако по мере накопления информации о пользователях и их предпочтениях, система получает возможность более точно настраивать свои алгоритмы и переходить к более сложным методам рекомендаций.

Модуль контентной фильтрации

При разработке модуля контентной фильтрации были созданы две ключевые функции `getCourseRecommendationBasedOnUserPerformance` и `getCourseRecommendationBasedOnUserProfile`.

Функция `getCourseRecommendationBasedOnUserPerformance` - рекомендует пользователю курсы на основе его оценок по предметам.

В первую очередь с помощью функции `createUserPerformanceSummary` составляется сводка успеваемости пользователя со средними оценками по каждому предмету. Затем система приводит значения сложности курсов к числовому представлению.

Для исключения не подходящих курсов на основе сводки успеваемости используется функция `deleteIrrelevantCoursesBasedOnUserPerformance`.

Функция удаляет курс из таблицы, если средняя оценка пользователя по хотя бы одному из предметов курса ниже минимально допустимого значения для соответствующего уровня сложности, определенного константами `UNDEFINED_THRESHOLD`, `BEGINNER_THRESHOLD`, `INTERMEDIATE_THRESHOLD`, `ADVANCED_THRESHOLD`. При этом `UNDEFINED_THRESHOLD` является особым случаем, при котором у пользователя нет оценок по предметам данного курса, в следствии чего такие курсы допустимы для рекомендации, но имеют низкий приоритет.

Перед тем, как методом косинусного сходства будет рассчитана степень схожести между данными о курсах и сводки успеваемости пользователя обе структуры должны быть приведены к единому виду:

1. Оценки в сводке успеваемости приводятся к значениям в интервале $[0, 1]$.
2. Каждый курс трансформируется в представление формата one-hot – encoding, где его связь с соответствующими предметами кодируются в виде бинарного вектора, отражающего наличие или отсутствие каждого из возможных предметов.

В завершении данного метода фильтрации система извлекает самые подходящие курсы, количество которых определяется переданным аргументом функции.

Функция «getCourseRecommendationBasedOnUserProfile» - рекомендует пользователю курсы на основе его поведенческого профиля.

Как и в предыдущей функции каждый курс приводится к бинарному вектору и с помощью метода косинусного сходства сравнивается с поведенческим профилем пользователя, чтобы в последствии сформировать рекомендацию.

Работа поведенческого профиля пользователя описана в следующей под-главе.

Поведенческий профиль пользователя

Фильтрация на основе контента рекомендует пользователю различные курсы, на которые стоит обратить внимание, используя для этого его поведенческий профиль.

Работа поведенческого профиля пользователя основана на неявных и явных взаимодействиях, которые мы рассматривали во второй главе магистерской диссертации.

При взаимодействии пользователя с курсом, в таблице «user_implicit_feedback» создаётся новая запись, в которой поле «event_type» отражает тип взаимодействия. Например, при нажатии на иконку курса

создаётся запись со значением «viewed», фиксируя минимальный уровень взаимодействия и первичный интерес пользователя. Если пользователь записывается на курс, создаётся ещё одна запись со значением «enrolled». Завершение курса приводит к созданию новой записи, где в поле «event_type» указывается значение «completed».

В случае, если пользователь ставит оценку курсу, в таблице «user_explicit_feedback» создаётся новая запись, в которой поле «rating» отражает оценку пользователем курса от 1 до 5.

Для реализации поведенческого профиля пользователя была создана функция «createUserProfile», которой передаются данные о неявных и явных взаимодействиях, а затем очищаются от записей сторонних пользователей, оставляя данные пользователя, взаимодействующего с системой. Также данные о неявных взаимодействиях сортируются по дате в убывающем порядке, а затем мы преобразуем каждый тип неявного взаимодействия в числовое представление.

Следующим шагом в создании поведенческого профиля пользователя является разделение данных на отдельные профили неявных и явных взаимодействий. Для этого мы создадим две функции: «createImplicitUserProfile» и «createExplicitUserProfile». Эти функции будут обрабатывать данные о неявных и явных взаимодействиях соответственно, создавая профили, которые можно использовать для дальнейшего анализа и рекомендаций.

Рассмотрим подробнее механизмы формирования профилей, начиная с обработки неявных данных:

- Инициализация профиля. Создается пустой профиль пользователя в виде таблицы, где каждая колонка соответствует предмету курса, а значения инициализируются нулями.
- Обработка взаимодействий. Система проверяет, какое количество неявных взаимодействий сделал пользователь, если количество взаимодействий больше чем 100, то тогда мы берем 100 последних.

В другом случае мы берем всё. Если пользователь не имеет никаких взаимодействий, то мы возвращаем пустой профиль. Когда система проходит через все записи, она обновляет профиль пользователя, добавляя «вес» данного взаимодействия ко всем предметам, связанным с курсом, с которым произошло взаимодействие.

- Нормализация профиля. После обработки все значения нормализуются с использованием функции «normalizeUserProfileMinMax», чтобы привести их к диапазону [0, 1].

Каждое значение в профиле неявных взаимодействий пользователя отражает интерес к тому или иному предмету.

Механизмы формирования профиля, основанного на явных данных, выглядит следующим образом:

- Инициализация профиля. Создается пустой профиль пользователя в виде таблицы, аналогично профилю для неявных данных. Также создается словарь «subjectRatingSummary», где ключ — название предмета, а значение — массив [сумма рейтингов, количество оценок].
- Обработка взаимодействий. Когда система проходит через все записи, она обновляет словарь, добавляя оценку пользователя к сумме рейтингов, а также увеличивает количество оценок на 1 ко всем предметам, связанными с курсом, с которым произошло взаимодействие.
- Расчет среднего рейтинга. Для каждого предмета в словаре рассчитывается средняя оценка на основе суммы рейтингов и количества оценок, а затем записывается в профиль пользователя.
- Нормализация профиля. После обработки все значения нормализуются с использованием функции «normalizeUserProfileMinMax», чтобы привести их к диапазону [0, 1].

Оба профиля (явный и неявный) объединяются в единый профиль с помощью метода взвешенной суммы. Вес для явного профиля задается

константой «EXPLICIT_PROFILE_WEIGHT», а для неявного — «IMPLICIT_PROFILE_WEIGHT». Это позволяет учитывать разные степени важности двух типов данных.

Модуль совместной фильтрации

При разработке модуля совместной фильтрации была создана ключевая функция «getCourseRecommendationBasedOnUserSimilarity». Данная функция рекомендует пользователю курсы на основе схожести предпочтений пользователей.

В первую очередь для генерации персонализированных рекомендаций в функции «createUserCourseRatingMatrix» создаётся матрица оценивания, где на основе таблицы всех явных взаимодействий сохраняются оценки каждого курса, поставленные всеми пользователями. Затем система вычисляет степень сходства текущего пользователя с другими пользователями, используя метод косинусного сходства на основе матрицы оценок, располагая их в порядке убывания.

После вычисления степени сходства система прогнозирует оценки курсов на основе оценок самых похожих пользователей. Для этого использовалась функция «predictCoursePreference» на каждом из курсов, который еще не был оценен текущим пользователем. Перед этим собираются все имеющиеся оценки пользователей по данному курсу. Исходя из них, система отбирает фиксированное количество пользователей с наибольшей степенью сходства, что позволяет учитывать только наиболее релевантных пользователей для формирования оценки курса. Под конец рассчитывается среднее арифметическое отобранных оценок пользователей с наибольшей степенью сходства.

Завершающим этапом формирования рекомендации является сортировка прогнозированных оценок курсов в порядке убывания для выявления наиболее предпочтительных курсов.

Модуль гибридной фильтрации

В этой под-главе мы представляем реализацию объединения различных методов фильтрации данных для достижения максимально персонализированного подхода. Контентная, совместная и не персонализированная фильтрации каждая по-своему полезна, но их сочетание позволяет значительно улучшить качество рекомендаций.

Используя методы, такие как каскадирование и взвешенное объединение, мы адаптируем фильтрации под уникальные предпочтения каждого пользователя. Такой подход позволяет системе гибко учитывать различные факторы и создавать более точные и персонализированные результаты, максимально соответствующие интересам и потребностям пользователя.

Генератор рекомендаций реализован классом под названием «RecommendationGenerator». Каждый объект класса предназначен для формирования рекомендаций определенному пользователю. При инициализации генератора рекомендаций система формирует объекты репозитории с последующим изъятием необходимых данных для рекомендации, таких как: образовательные мероприятия, не завершенные пользователем, явные и неявные взаимодействия, а также оценки успеваемости пользователя по предметам.

Рассмотрим реализацию основных методов класса «RecommendationGenerator».

Первым методом является «generateRecommendationWeighted», который использует метод «взвешенного объединения», при котором ранги образовательных мероприятий, полученные в результате обработки совместной и контентной фильтрациями, совмещаются с помощью линейной комбинации в общий ранг. При этом в системе совместная фильтрация, является более важной, чем контентная, так как она ориентируется на прямые потребности пользователя, поэтому были определены «веса» для каждой из фильтраций, где совместная фильтрация получила больший коэффициент, равный 0,6. Контентная же в свою очередь получила коэффициент равный –

0,4. На основе общих рангов образовательные мероприятия сортируются в убывающем порядке для отображения наиболее предпочтительных рекомендаций.

Вторым методом является «generateRecommendationCascading», который использует метод «каскадирования» для формирования рекомендаций на основе сводки успеваемости, поведенческого профиля пользователя и матрицы оценок курсов. Сначала из всех не пройденных образовательных мероприятий пользователя отбираются 50 наиболее релевантных, используя функцию «getCourseRecommendationBasedOnUserPerformance», описанную в разделе «Модуль контентной фильтрации». Затем из выбранных 50 мероприятий исключаются 20 наименее подходящих, оцениваемых на основании поведенческого профиля пользователя, включающего такие параметры, как интересы и предыдущий опыт. На следующем этапе для оставшихся 30 мероприятий применяется совместная фильтрация. В результате отбираются 10 мероприятий, максимально удовлетворяющих требованиям и интересам пользователя.

Третьим и ключевым методом, участвующим в генерации рекомендаций, является «generateRecommendationSwitching», который представляет условные сценарии рекомендаций. Чтобы определить тот или иной сценарий используются три условия:

1. Достаточное количество неявных взаимодействий с системой, чей минимальный порог определяется константой «MINIMUM_NUMBER_OF_IMPLICIT_INTERACTIONS».
2. Достаточное количество неявных взаимодействий с системой, чей минимальный порог определяется константой «MINIMUM_NUMBER_OF_EXPLICIT_INTERACTIONS».
3. Достаточное количество оценок по предметам «MINIMUM_NUMBER_OF_GRADES».

Данные условия обоснованы тем, что при недостаточном количестве данных, полученных от пользователя, поведенческий профиль пользователя, матрицы оценок курсов и сводки успеваемости пользователя теряют свою актуальность для формирования рекомендаций, поскольку объем предоставленных данных недостаточен для эффективного применения в фильтрациях, направленных на реализацию персонализированного подхода.

Всего в данном методе используется 8 сценариев.

Первый сценарий запускается в случае, если у пользователя не имеется достаточного количества оценок по предметам, но имеется достаточное количество явных и неявных взаимодействий. В этом случае применяется взвешенный гибридный подход, который объединяет совместную фильтрацию и рекомендации на основе профиля пользователя.

Второй сценарий запускается в случае, если в первом сценарии не было выполнено условие достаточного количества явных взаимодействий. В таком случае используются рекомендации на основе профиля пользователя, которые формируются на основе неявной обратной связи.

Третий сценарий запускается в случае, если в первом сценарии не было выполнено условие достаточного количества неявных взаимодействий. В таком случае применяется подход, основанный на сходстве с другими пользователями. Для этого используются явная обратная связь и информация о предпочтениях схожих пользователей.

Четвертый сценарий запускается в случае «холодного» старта системы. В таком случае формируются рекомендации на основе самых популярных курсов среди всех пользователей.

Пятый сценарий запускается в случае если у пользователя имеются все необходимые данные. В таком случае применяется метод каскадирования, который объединяет: данные об успеваемости, поведенческий профиль пользователя и совместную фильтрацию.

Шестой сценарий запускается в случае, если в пятом сценарии не было выполнено условие достаточного количества явных взаимодействий. В таком

случае применяется метод каскадирования, который объединяет данные об успеваемости и поведенческий профиль пользователя.

Седьмой сценарий запускается в случае, если в пятом сценарии не выполнено условие достаточного количества неявных взаимодействий. В таком случае применяется метод каскадирования, который объединяет данные об успеваемости и совместную фильтрацию.

Восьмой сценарий запускается в случае, если у пользователя есть только достаточное количество оценок по предметам. В этом случае система формирует рекомендацию, основываясь только на сводке успеваемости пользователя.

Таким образом в данном разделе подробно рассматривалась техническая реализация модулей не персонализированной, контентной и совместной фильтраций, а также их объединения в модуле гибридной фильтрации.

Определение метрик для оценки эффективности рекомендаций

Качество прогнозирования в рекомендательных системах оценивают с помощью регрессионных, классификационных, контекстуальных метрик, а также метрик ранжирования. Выбор подходящих метрик зависит от особенностей данных и целей, которые должна решать рекомендательная система.

Регрессионные метрики предназначены для оценки точности прогнозов, выраженных в числовой форме. Это может помочь определить, насколько модель близка к фактическим значениям и насколько точно она предсказывает результаты, такие как оценки образовательных мероприятий и вероятность успеха в освоении образовательного мероприятия. Данная метрика не подходит для тестирования эффективности выдаваемых рекомендаций предлагаемой гибридной рекомендательной системы по причине того, что образовательные мероприятия, использующиеся в системе, не выражаются в числовой форме.

Метрики ранжирования оценивают порядок рекомендаций в списке, проверяя, насколько высоко располагаются наиболее полезные элементы. Они

играют важную роль, поскольку пользователи чаще всего обращают внимание на верхние позиции в списке рекомендаций. Данные метрики подходят для тестирования эффективности предлагаемой гибридной рекомендательной системы, так как важно не только определить релевантные образовательные мероприятия, но и расположить их в таком порядке, чтобы наиболее подходящие варианты отображались первыми.

Контекстуальные метрики измеряют качество рекомендаций с точки зрения их уникальности, новизны и разнообразия. Они фокусируются на пользовательском опыте, оценивая, насколько рекомендации соответствуют интересам пользователя и выдают полезные рекомендации, которые не ожидает пользователь. Данная метрика не подходит для тестирования эффективности выдаваемых рекомендаций предлагаемой гибридной рекомендательной системы по причине того, что реализуемая система опирается на предоставление образовательных мероприятий максимально подходящих под актуальные потребности пользователя не учитывая уникальность и разнообразие.

Классификационные метрики используются для оценки способности системы правильно определять релевантность объектов в соответствии с предпочтениями пользователя. Эти метрики помогают измерить, насколько эффективно система классифицирует элементы как подходящие или неподходящие для конкретного пользователя, основываясь на его предпочтениях и потребностях. Данная метрика подходит для тестирования эффективности выдаваемых рекомендаций предлагаемой гибридной рекомендательной системы по причине того, что позволяют понять, насколько хорошо система удовлетворяет основные образовательные потребности пользователей.

Для определения качества прогнозирования предлагаемой гибридной рекомендательной системы было решено рассмотреть наиболее популярные классификационные метрики и метрики ранжирования, такие как:

- точность на K элементах (precision@k);

- полнота на K элементах (recall@k);
- среднее гармоническое точности и полноты при K (F-score@k);
- вероятность попадания подходящего элемента (Hit Rate);
- средний обратный ранг (Mean Reciprocal Rank).

Вероятность попадания подходящего элемента (Hit Rate) – это метрика, принимающая бинарные значения, которая указывает, был ли хотя бы один подходящий элемент среди первых k предложений. Если хотя бы один из них соответствует ожиданиям, Hit Rate принимает значение 1, в противном случае - 0.

Вероятность попадания подходящего элемента можно представить в виде следующей формулы (22):

$$HR = \frac{\sum_{v=1}^N relevance_v}{N}, \quad (22)$$

где $relevance_v$ – {1, если в рекомендации для пользователя v был релевантный объект, в противном случае 0};

N – общее количество пользователей;

m – рекомендуемое образовательное мероприятие.

Средний обратный ранг (Mean Reciprocal Rank) – это среднее значение обратных рангов для первого подходящего элемента в списке рекомендаций. MRR оценивает, на каком месте в ранжированном списке появляется первый подходящий элемент, и служит для анализа качества ранжирования в поисковых и рекомендательных системах.

Средний обратный ранг можно представить в виде следующей формулы (23):

$$MRR = \frac{\sum_{v=1}^N RR_v}{N} \quad RR_v = \frac{relevance_v}{rank_v}, \quad (23)$$

где $rank_v$ – номер первого подходящего образовательного мероприятия в рекомендации для пользователя v.

Точность на K элементах ($\text{Precision@}k$) – это метрика, оценивающая, сколько из первых k рекомендованных объектов являются подходящими. Она отражает степень точности системы в представлении пользователю наиболее подходящих элементов среди первых k предложений (24).

$$\text{Precision@}K = \frac{\sum_{m \leq K} \text{relevance}_{vm}}{K}, \quad (24)$$

где K – количество образовательных мероприятий, рекомендуемых пользователю.

Полнота при отсечении K ($\text{Recall@}k$) – это метрика, которая оценивает, какую часть всех подходящих элементов система включает в свои первые k рекомендаций. Она показывает, насколько эффективно система покрывает все подходящие объекты, ограничиваясь лишь первыми k предложениями (25).

$$\text{Recall@}K = \frac{\sum_{m \leq K} \text{relevance}_{vm}}{L}, \quad (25)$$

где L – число объектов в отложенной выборке, с которыми взаимодействовал пользователь.

Буква K в названии метрик обозначает количество рекомендованных элементов, которые мы учитываем в анализе. Обычно K выбирают из стандартных значений, таких как 5, 10, 20 или 50, однако точное число зависит от специфики задачи. Например, это может быть количество образовательных мероприятий, отображаемых на первой странице. Стоит отметить, что с увеличением K метрика $\text{recall@}k$, как правило, возрастает, потому что с большим количеством предложений возрастает вероятность того, что система захватит больше релевантных объектов. В то же время $\text{precision@}k$ может снизиться, так как с расширением списка рекомендаций растет вероятность того, что менее релевантные элементы попадут в топ.

Среднее гармоническое точности и полноты при K ($F\text{-score}@k$) – это среднее значение, вычисляемое как гармоническая средняя между точностью на K элементах ($Precision@k$) и полнотой при отсечении K ($Recall@k$), что позволяет учитывать оба аспекта качества рекомендаций в одной метрике. $F\text{-score}@k$ применяется для достижения баланса между $Precision@k$ и $Recall@k$ в пределах первых k рекомендаций.

Среднее гармоническое точности и полноты при K ($F\text{-score}@k$) можно представить в виде следующей формулы (26):

$$F_{\beta} = \frac{(1 + \beta^2) * Precision@K * Recall@K}{(\beta^2 * Precision@K) + Recall@K} \quad (26)$$

Так как точность и полнота важны в равной мере, используется значение 1 для β .

Таким образом в данном разделе были определены и описаны метрики тестирования предлагаемой гибридной рекомендательной системы.

Вывод по третьей главе

Основывая на теоретической информации, полученной в главах 1 и 2 магистерской диссертации, была спроектирована и описана гибридная рекомендательная система анализирующая успеваемость школьников на основании данных об оценках для персонализированной выдачи рекомендаций по обучающим мероприятиям. Также были описаны метрики тестирования, необходимые для оценки эффективности выдаваемых рекомендаций предлагаемой гибридной рекомендательной системой.

4 Тестирование предлагаемой гибридной рекомендательной системы

4.1 Проверка точности прогнозирования интересов пользователя в гибридной рекомендательной системе

Оценка эффективности рекомендательных систем — это ключевая задача, которая позволяет понять, насколько успешно система удовлетворяет потребности пользователей, предлагая им подходящие и полезные рекомендации. В данном разделе будет описан алгоритм, используемый для оценки эффективности гибридной рекомендательной системы, которая применяется для рекомендаций образовательных мероприятий. Алгоритм включает несколько шагов, от подготовки данных до расчета финальных метрик, которые позволят оценить качество работы системы.

Во время разработки алгоритмов тестирования эффективности рекомендации были выделены следующие шаги:

1. Подготовка тестовых данных. Перед оценкой эффективности системы необходимо подготовить и загрузить тестовые данные, которые будут использоваться для проверки её работоспособности. Такие данные включают в себя: информацию об образовательных мероприятиях, тестовых пользователях, их явные и не явные взаимодействия, а также их оценки по предметам. Для обеспечения наглядности тестирования данные должны быть достаточно разнообразными и отражать интересы пользователей и их поведение. Также важно, чтобы данные о предпочтениях пользователей были актуальными и содержали достаточное количество информации для того, чтобы система могла сделать точные предсказания.
2. Построение матрицы сходства. Одним из важных этапов является создание матрицы сходства, которая используется для оценки схожести между различными образовательными мероприятиями. В

нашем случае для построения матрицы сходства используется метод TF-IDF (Term Frequency-Inverse Document Frequency) для векторизации текстового контента мероприятий. Это позволяет преобразовать описание, название и связанные предметы каждого мероприятия в числовое представление, которое далее используется для вычисления косинусного сходства между мероприятиями. Косинусное сходство является эффективным методом для измерения схожести между векторами векторизованных текстов и используется, как инструмент для определения актуальности рекомендованного образовательного мероприятия в соответствии с предпочтениями пользователя.

3. Формирование групп тестовых пользователей. Перед проведением тестирования системы необходимо разделить пользователей на несколько групп, для получения более наглядных и точных результатов тестирования. Это обусловлено тем, что в реальных условиях могут возникать значительные различия в результатах между подгруппами. Например, одна группа может демонстрировать отличные результаты, тогда как другая — наоборот, очень слабые. Если такие данные рассматривать в общем виде, усредненные метрики могут выглядеть весьма положительно, но они не дадут полной картины происходящего и могут скрыть реальные проблемы и недостатки в работе системы. Разделение на группы позволяет выявить такие различия и глубже проанализировать эффективность решения для каждой подгруппы. В данном случае тестовые пользователи разделяются на группы по пять человек в каждой. Такой размер групп позволяет учесть разнообразие предпочтений и поведения пользователей, а также избежать искажения результатов, которое может быть связано со слишком малым числом участников.
4. Расчёт метрик оценки эффективности. Для каждого пользователя в группе рассчитываются следующие метрики: precision@k , recall@k ,

F-score. В рамках нашего эксперимента значение K было равно 20. Далее определяется среднее арифметическое для каждой из вышеупомянутых метрик в контексте одной группы. Финальная оценка эффективности системы рассчитывается путём нахождения среднего значения для каждой из метрик на основе полученных средних значений каждой группы. Это позволяет интегрировать результаты всех групп и получить более общую картину работы системы. Также этот шаг включает в себя расчет метрик HR (Hit Rate) и MRR (Mean Reciprocal Rank) на основе всех пользователей независимо от группы.

Далее будет представлена подробная интерпретация полученных данных, включая визуализацию метрик, что позволит лучше понять поведение системы в реальных условиях. Эти результаты помогут оценить, насколько точно система может рекомендовать подходящие образовательные мероприятия, а также позволят сравнить эффективность различных подходов и выявить возможные слабые места в текущем решении.

Оценка отношения пользователей, которые получили хотя бы одну подходящую рекомендацию к общему количеству пользователей представлена метрикой HR и составляет 1.00.

Значение HR (Hit Rate), равное 1.00, указывает на то, что все пользователи получили хотя бы одну релевантную рекомендацию, что подтверждает полноту системы и её способность находить подходящие образовательные мероприятия для каждого пользователя.

Оценка усредненной позиции первого подходящего образовательного мероприятия в списке рекомендаций среди всех пользователей представлена метрикой MRR и составляет 0.86.

Значение MRR (Mean Reciprocal Rank), равное 0.86, свидетельствует о том, что в среднем первый релевантный элемент находится на высокой позиции в списке рекомендаций. Это означает, что система эффективно

выстраивает порядок рекомендаций, предоставляя пользователям наиболее подходящие мероприятия в начале списка.

В таблице 2 представлены результаты тестирования с помощью метрик: precision@k, recall@k и F-score.

Таблица 2 – Результаты эксперимента

№ группы	Точность первых 20 элементов	Полнота первых 20 элементов	F-мера
1	0,75	0,84	0,78
2	0,76	0,91	0,81
3	0,77	0,75	0,75
4	0,7	0,71	0,67
5	0,7	0,78	0,72
6	0,91	0,87	0,89
7	0,77	0,51	0,59
8	0,9	0,77	0,82
Среднее значение	0,82	0,77	0,76

На рисунке 13 изображен график, основывающийся на данных, полученных по результатам экспериментов.

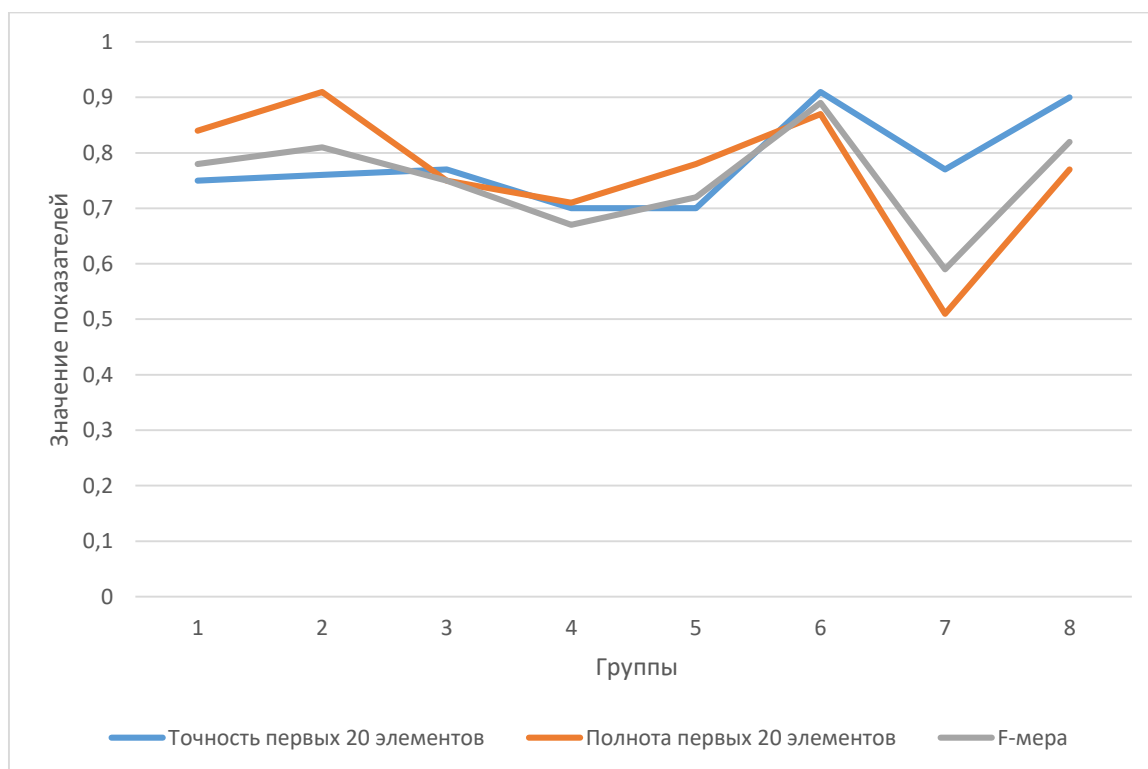


Рисунок 13 – График экспериментальных исследований

Средние значения precision@20 (0,82), recall@20 (0,77) и F1-score (0,76) демонстрируют высокие показатели, что указывает на высокое качество рекомендаций в большинстве случаев. Однако группа 7 выделяется низкими значениями по всем метрикам, особенно по recall@20 (0,51).

Результаты описанные выше показывают, что рекомендательная система показывает хорошие результаты в решении поставленной задачи.

В данном разделе были описаны результаты проведенных тестов предлагаемой гибридной рекомендательной системы.

4.2 Тестирование способности интеграции гибридной рекомендательной системы в веб-приложения

Для тестирования интеграции гибридной рекомендательной системы с веб-приложениями был создан прототип образовательной платформы, который использовался как тестовая среда для проверки предложенного подхода к прогнозированию интересов пользователей.

При реализации платформы использовался веб-фреймворк Django, поскольку он работает на языке программирования Python, как и разрабатываемая гибридная рекомендательная система, что упрощает интеграцию, обеспечивает единый стек технологий и ускоряет разработку.

Интеграция с системой происходит при помощи серверного рендеринга страниц (Server-Side Rendering, SSR) с использованием Django. В этом подходе веб-приложение взаимодействует с сервером через HTTP-запросы. Когда пользователь открывает страницу, браузер отправляет GET-запрос к серверу, который обрабатывается Django. Далее Django получает необходимые данные, подставляет их в HTML-шаблон с помощью встроенного шаблонизатора и отправляет браузеру уже полностью сформированную страницу.

Визуальная часть веб-приложения была реализована на языке гипертекстовой разметки – HTML, языке описания внешнего вида – CSS.

Хранение данных осуществлялось благодаря свободной объектно-реляционной системе управления базами данных PostgreSQL, по причине ее популярности среди веб-приложений. Для отслеживания состояния базы данных использовался веб-интерфейс для администрирования – pgAdmin.

С помощью Docker была создана изолированная среда в которой хранилась база данных для осуществления её независимой работы от основной системы.

Данные успеваемости пользователей были получены из МБУ «Школа №16».

Тестовые данные об образовательных мероприятиях были взяты с онлайн-платформы для соревнований по анализу данных и машинному обучению Kaggle.

При первом посещении платформы пользователю необходимо совершить регистрацию в системе для создания учетной записи, которая необходима для последующего отслеживания активности пользователя и его оценок. После авторизации, пользователь попадает на главную страницу, где для него доступна вкладка «Курсы», перейдя на которую система формирует рекомендацию и выдаёт страницу с предлагаемыми образовательными мероприятиями.

При первом переходе на страницу с образовательными мероприятиями у системы не имеется достаточного количества данных, для выдачи персонализированных мероприятий, в связи с чем система формирует не персонализированную рекомендацию на основе общей популярности образовательных мероприятий.

Наиболее популярные образовательные мероприятия, среди пользователей образовательной платформы, представлены на рисунке 14.

Здравствуйте, ваши рекомендации Иван Иванов!

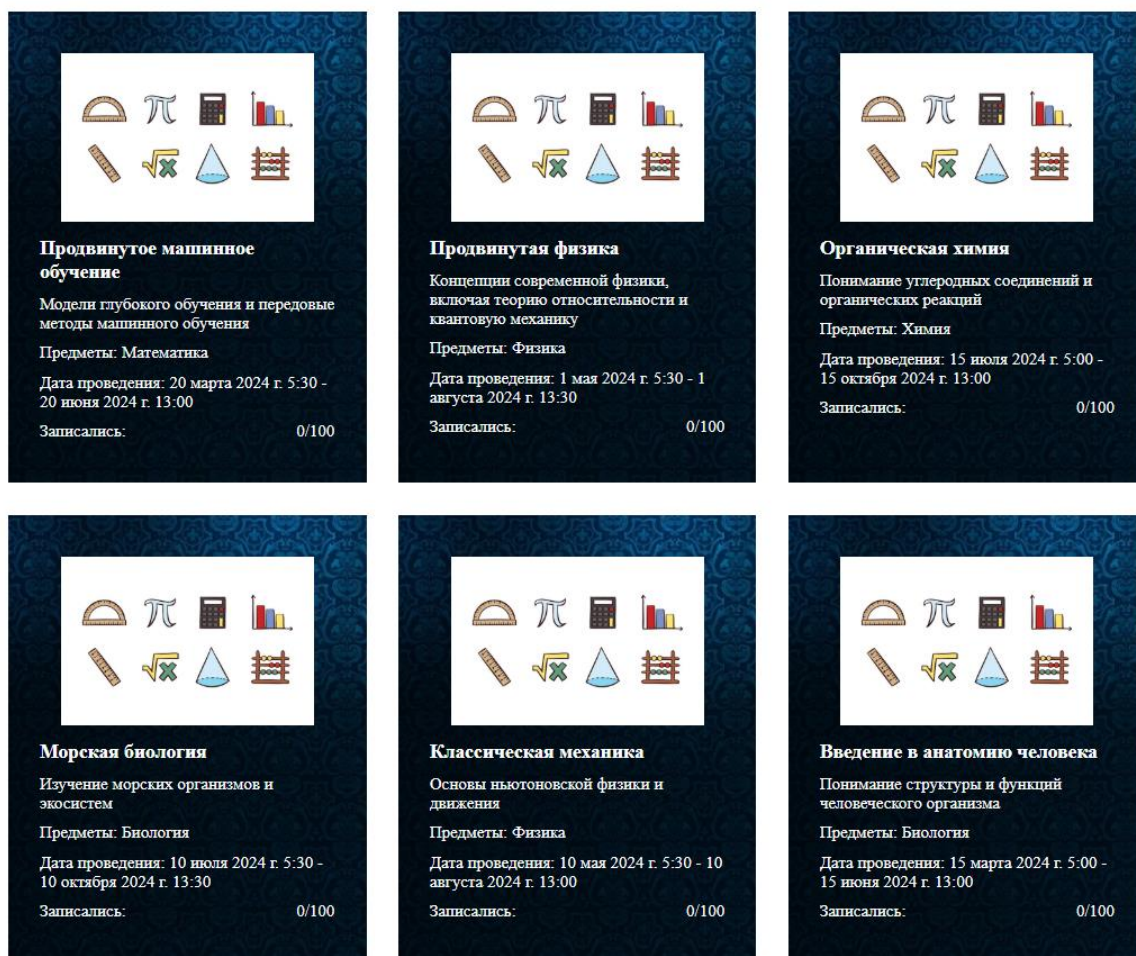


Рисунок 14 – Наиболее популярные образовательные мероприятия

На рисунке 15 представлена таблица из базы данных отображающая количество посещений наиболее популярных образовательных мероприятий.

	course_id [PK] integer	name character varying	number_of_visits integer
1	63	Advanced Machine Learning	11500
2	85	Advanced Physics	11000
3	107	Organic Chemistry	9900
4	95	Classical Mechanics	9500
5	113	Introduction to Human Anatomy	9500
6	37	Marine Biology	9500

Рисунок 15 – Количество посещений наиболее популярных образовательных мероприятий

При нажатии на карточку образовательного мероприятия пользователь переходит на страницу с подробным описанием информации о нём. В момент перехода система фиксирует посещение курса, создавая в базе данных событие неявного взаимодействия со значением «viewed».

На рисунке 16 представлен пример страницы с подробным описанием образовательного мероприятия на тему «Продвинутая география».



Рисунок 16 – Страница с подробным описанием образовательного мероприятия

Там же пользователь может записаться на курс, завершить его и поставить оценку.

При нажатии на кнопку «Записаться» или «Завершить» система фиксирует неявное взаимодействие пользователя, создавая в базе данных соответствующее событие со значением «enrolled» или «completed».

На рисунке 17 представлен сценарий посещения, записи и завершения образовательного мероприятия.

	interaction_id [PK] integer	user_id integer	course_id integer	event_type character varying	timestamp timestamp without time zone
1	5	1	63	viewed	2024-12-21 19:05:49.766353
2	6	1	63	enrolled	2024-12-21 19:16:56.249392
3	7	1	63	completed	2024-12-21 19:16:57.549137

Рисунок 17 – Неявные взаимодействия пользователя с системой, сохраненные в базе данных

В случае, если пользователю понравится или не понравится пройденное образовательное мероприятие, на странице с подробным описанием курса есть возможность поставить оценку. Данная оценка также фиксируется системой в базе данных, как явное взаимодействие пользователя с системой.

Далее на основе примеров будут описаны алгоритмы и их условия при которых они могут работать, как отдельности друг от друга, так и совместно, используя методы гибридных рекомендательных систем. В процессе тестирования в первую очередь были проведены тесты каждого алгоритма по отдельности, а затем были совмещены все вместе.

Получив достаточное количество неявных взаимодействий, система применяет контентную фильтрацию на основе поведенческого профиля пользователя, представленного в таблице 3.

Таблица 3 – Поведенческий профиль пользователя

Биология	0,15
Химия	0,15
География	0,3
Математика	0
Философия	0
История	0,075
Физика	0
ИЗО	0,15
Английский язык	0
Музыка	0,075

Если у пользователя имеются явные взаимодействия, то они также используются в процессе формирования профиля пользователя.

Поскольку в поведенческом профиле пользователя основной интерес направлен на географию, система в первую очередь рекомендует образовательные мероприятия, связанные с этим предметом. При этом, если возможно, она дополнительно учитывает мероприятия, где вторым или последующим предметом выступает другой предмет из профиля пользователя с коэффициентом ниже, чем у предмета с самым высоким коэффициентом.

На рисунке 18 представлен результат контентной фильтрации основанной на поведенческом профиле пользователя.

Здравствуйте, ваши рекомендации Иван Иванов!

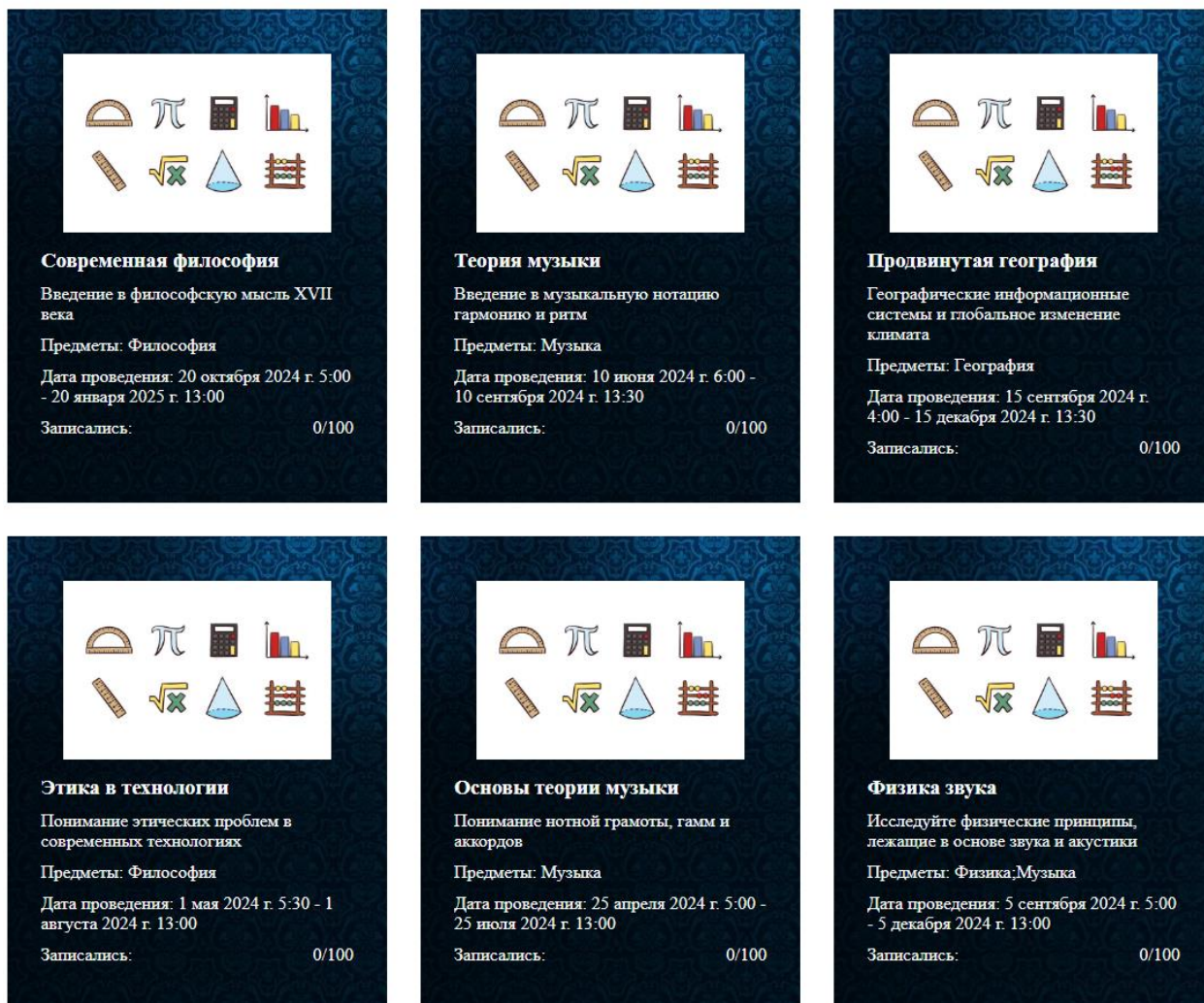


Рисунок 18 – Результат контентной фильтрации, основанный на поведенческом профиле конкретного пользователя

В случае, если у пользователя проставлено достаточное количество оценок по предметам, система формирует рекомендацию используя его сводку успеваемости, представленную в таблице 4.

Таблица 4 – Средние оценки пользователя по предметам

Биология	4,8
Химия	5
География	4,5
Математика	2,5
Философия	4,3
История	2,4
Физика	2,6
ИЗО	4,5
Английский язык	2,6
Музыка	2,7

Так как в сводке успеваемости пользователя наивысшее среднее арифметическое оценок принадлежит химии, система в первую очередь предлагает образовательные мероприятия, связанные с этим предметом. При этом, если возможно, она также учитывает мероприятия, где вторым или последующим предметом выступает другой предмет из сводки успеваемости со средним арифметическим, уступающим максимальному значению.

На рисунке 19 представлен результат контентной фильтрации на основе сводки успеваемости пользователя.

Здравствуйте, ваши рекомендации Иван Иванов!

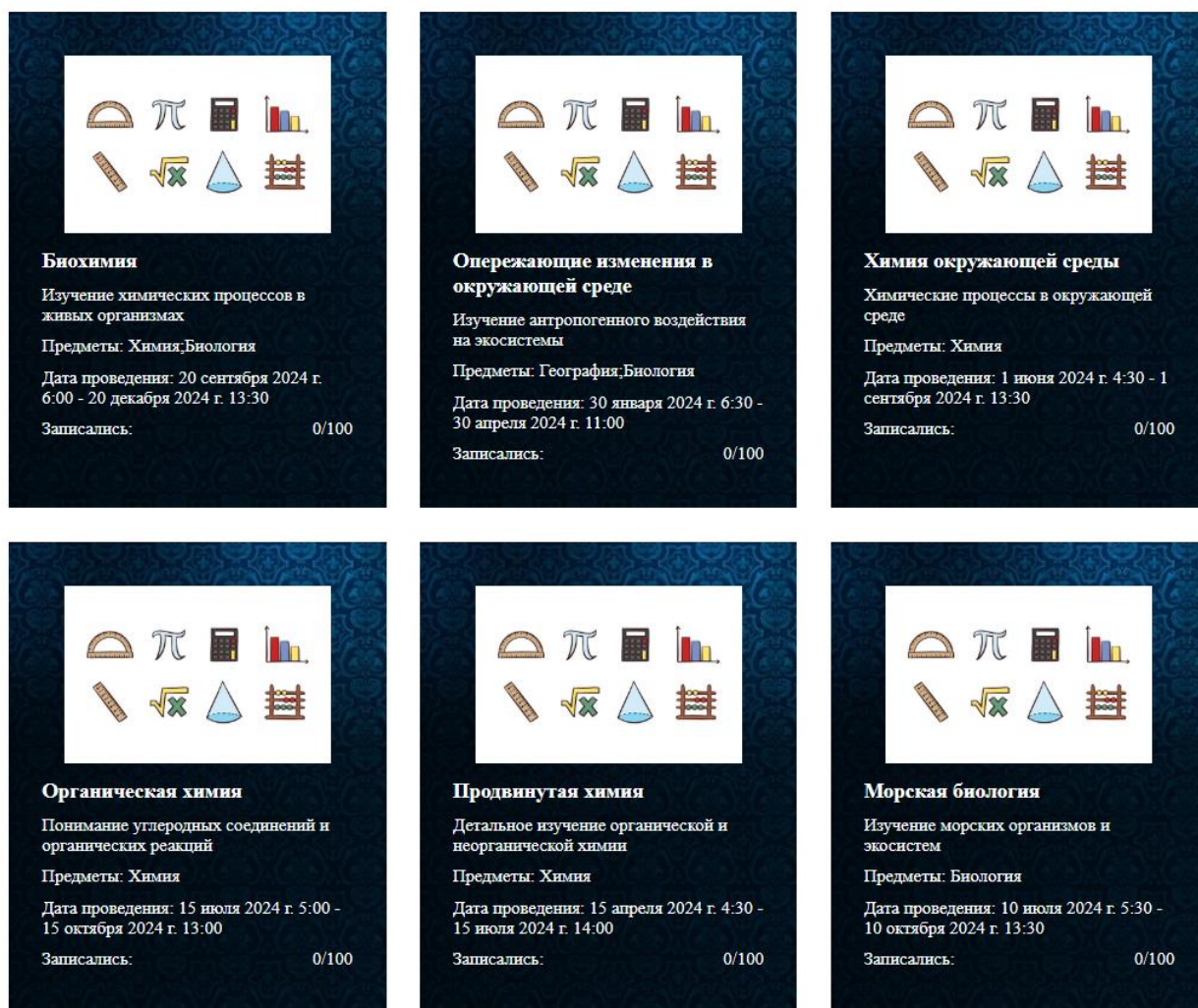


Рисунок 19 - Результат контентной фильтрации на основе успеваемости

Когда система накапливает достаточное количество явных взаимодействий от пользователей, например, выставленных оценок образовательным мероприятиям, она переходит к использованию методов коллаборативной фильтрации. Этот подход позволяет строить рекомендации, опираясь на поведенческие паттерны схожих пользователей, предполагая, что пользователи с похожими интересами будут положительно реагировать на одни и те же объекты. В таблице 5 представлены спрогнозированные оценки рекомендованных образовательных мероприятий на основе оценок самых похожих пользователей. Представленные данные наглядно демонстрируют, как коллаборативная фильтрация позволяет предсказывать интересы текущего

пользователя, даже если он ранее не взаимодействовал с конкретными мероприятиями.

Таблица 5 – Спрогнозированные оценки рекомендованных образовательных мероприятий

Имя курса	Спрогнозированная оценка	Оценки самых похожих пользователей
Основы художественной критики	5.0	[5.0]
Искусство кино	5.0	[5.0]
Математическая логика	5.0	[5.0]
Эпоха Возрождения и ее влияние	5.0	[5.0]
Английская литература	5.0	[5.0]
Бизнес-стратегия	5.0	[5.0]

Рекомендация образовательных мероприятий, представленных в таблице, основана на высокой спрогнозированной оценке, которая определяется как среднее значение оценок этого образовательного мероприятия от пользователей, имеющих схожие предпочтения.

Рекомендации образовательных мероприятий, представленных в таблице 5, сформированы на основе метода коллаборативной фильтрации, использующего оценки пользователей с наиболее схожими интересами. Каждому мероприятию присвоена спрогнозированная оценка, которая представляет собой среднее значение реальных оценок, выставленных другими пользователями, чьи поведение и интересы наиболее близки к текущему пользователю. Как видно из таблицы, все курсы получили максимальную спрогнозированную оценку — 5.0, что свидетельствует о высокой вероятности того, что текущий пользователь также положительно воспримет данные мероприятия. Это говорит о сильной корреляции между предпочтениями пользователя и уже имеющимися оценками у похожих участников системы.

На рисунке 20 представлен результат коллаборативной фильтрации.

Здравствуйте, ваши рекомендации Иван Иванов!

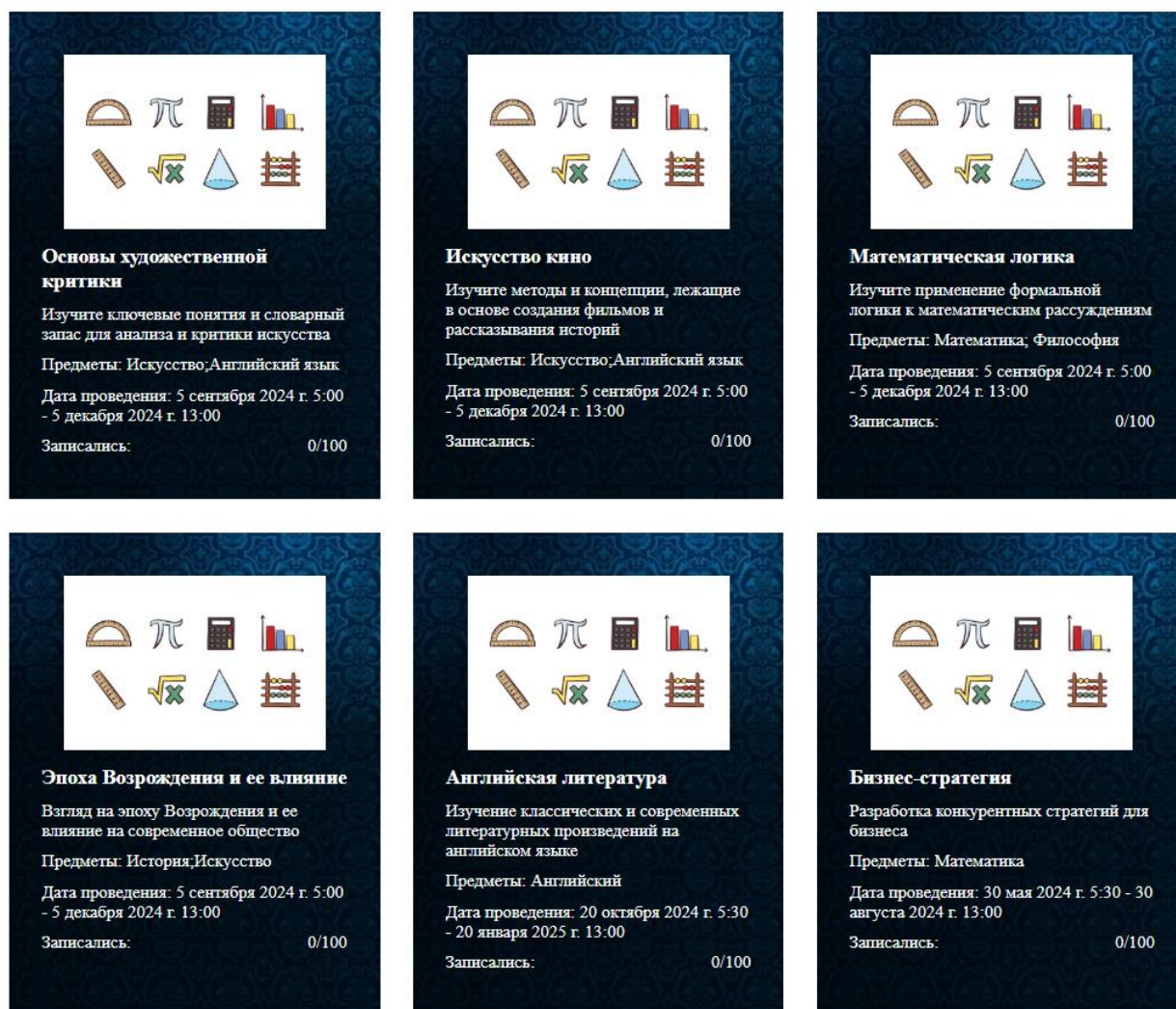


Рисунок 20 - Результат коллаборативной фильтрации

Была рассмотрена работа не персонализированной, контентной и коллаборативной фильтраций, проиллюстрировано их применение в предлагаемой гибридной рекомендательной системе. Теперь, на основе всех трёх методов, мы можем увидеть результаты работы системы, представленные в виде рисунка 21, который демонстрирует, как система комбинирует информацию из разных фильтраций для формирования наиболее подходящих рекомендаций.

Здравствуйте, ваши рекомендации Иван Иванов!

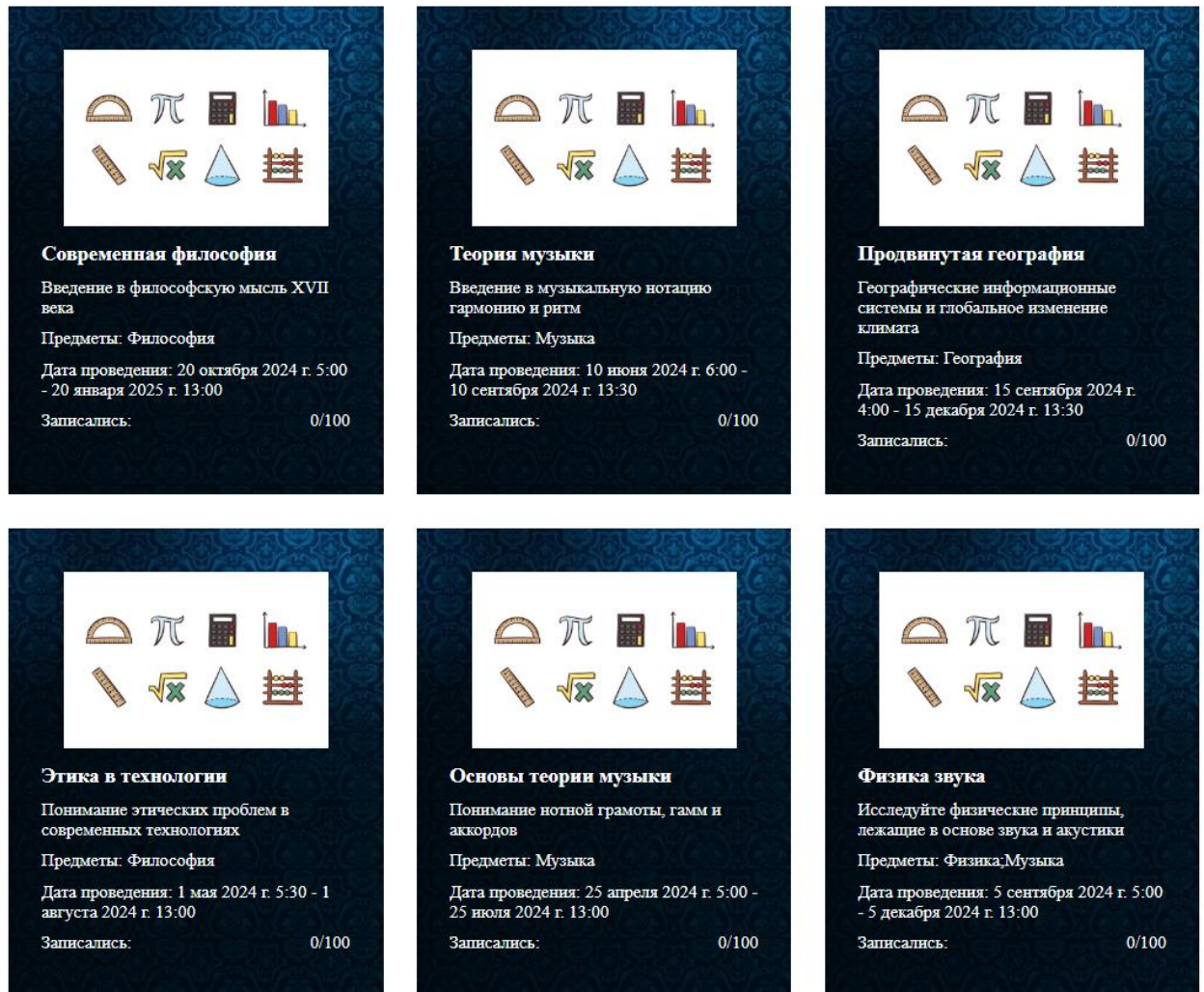


Рисунок 21 - Результат гибридной фильтрации

В данном разделе была описана успешная попытка интеграции, предлагаемой гибридной рекомендательной системы в веб-приложение.

4.3 Анализ результатов проведенного тестирования предлагаемой гибридной рекомендательной системы

В ходе тестирования предлагаемой гибридной рекомендательной системы были проанализированы различные метрики качества рекомендаций, такие как precision@k , recall@k , F1-score, HR (Hit Rate) и MRR (Mean Reciprocal Rank). Результаты показали, что система демонстрирует хорошие показатели точности и полноты рекомендаций, однако также выявились

несколько областей, требующих внимания для улучшения качества рекомендаций.

Средние значения по precision@k , recall@k и F1-score составили 0.82, 0.77 и 0.76 соответственно. Это указывает на то, что система эффективно генерирует релевантные рекомендации, однако существует некоторое пространство для улучшения, особенно в части повышения полноты рекомендаций, что может быть связано с ограниченным количеством подходящих курсов для некоторых пользователей.

Особенно низкие значения были зафиксированы для группы 7, которая продемонстрировала минимальные результаты по recall@k и F1-score (0.51 и 0.59 соответственно). Это может свидетельствовать о проблемах с персонализацией рекомендаций и недостаточным покрытием интересов пользователей этой группы, что требует дополнительной настройки фильтрации и улучшения алгоритмов под определенные пользовательские сегменты.

Показатели HR (1.00) и MRR (0.86) продемонстрировали высокую эффективность системы в нахождении хотя бы одной релевантной рекомендации и в обеспечении высокого положения первого релевантного элемента в списке. Это говорит о том, что система правильно ранжирует рекомендованные мероприятия, начиная с наиболее подходящих.

Несмотря на положительные результаты в части HR и MRR, система могла бы столкнуться с проблемами холодного старта для новых пользователей или мероприятий, в случае малого количества данных для точной персонализированной рекомендации, что могло бы ограничивать возможности системы при рекомендациях для новых пользователей. Данная проблема была решена внедрением модуля не персонализированной фильтрации.

После анализа результатов тестирования с использованием стандартных метрик качества, были проведены пользовательские тесты в специально созданной тестовой среде в виде веб-сайта. В ходе этих тестов было выявлено

несколько ключевых проблем в работе, предложенной гибридной рекомендательной системы.

Одной из сильных сторон предлагаемой гибридной рекомендательной системы была выявлена её способность учитывать успеваемость ученика при формировании рекомендаций, что позволяет адаптировать предложенные образовательные мероприятия под уровень знаний и подготовки каждого пользователя. Система анализирует текущие оценки и успеваемость ученика, что помогает ей отсеивать потенциально сложные образовательные мероприятия, которые могут быть слишком трудными или неподходящими для его текущего уровня знаний. Это важный аспект, поскольку он позволяет избежать перегрузки ученика материалом, который он может не в состоянии усвоить на данном этапе обучения.

В контексте коллаборативной фильтрации, когда система предсказывает оценку пользователя на основе оценок похожих пользователей, была обнаружена проблема разреженности данных, что для большинства курсов пользователи оставляют оценки только для ограниченного числа элементов. Это приводит к тому, что для большинства пользователей нет достаточного количества общих оценок с другими пользователями, чтобы вычислить сходство между ними. В результате система сталкивается с трудностью в нахождении похожих пользователей по методу оценивания, поскольку для точного расчёта сходства требуется достаточно оценок, чтобы выявить паттерны предпочтений. Когда большинство пользователей оценивают лишь небольшую часть курсов, предсказания становятся менее точными, а точность рекомендаций значительно снижается. Данную проблему решает модуль контентной фильтрации.

Одной из потенциальных проблем, выявленных в ходе тестирования, является отсутствие разнообразия в рекомендациях. Система предлагает пользователю курсы исключительно в рамках узкой области его интересов, такие как математика, что ограничивает его выбор и не позволяет раскрыть потенциал для предложения более широкого спектра образовательных

мероприятий. Данная проблема может привести к тому, что пользователи не получают возможность узнать о других курсах, которые могли бы им понравиться, но которые не попадают в узкий круг их интересов.

В данном разделе был проведён анализ полученных результатов тестирования предлагаемой гибридной рекомендательной системы.

4.4 Будущее развитие предлагаемой гибридной рекомендательной системы

Улучшение алгоритмов обработки данных

Одним из основных направлений совершенствования является разработка и улучшение алгоритмов для анализа и обработки данных, среди которых можно выделить:

- Применение продвинутых методов векторизации. Вместо традиционных подходов, таких как TF-IDF, можно использовать более современные технологии, такие как word2vec, GloVe или BERT. Эти модели способны лучше улавливать семантические связи между словами, что способствует повышению точности матриц сходства и, как следствие, улучшению качества рекомендаций.
- Объединение данных из разнообразных источников. Будущие усовершенствования могут включать интеграцию новых типов данных, таких как отзывы пользователей, информация из социальных сетей и поведенческие метрики (например, клики, время взаимодействия с контентом, просмотры). Это позволит более глубоко понимать предпочтения пользователей и предлагать им максимально подходящие рекомендации.

Персонализация рекомендаций

Ключевым шагом на пути развития гибридной рекомендательной системы является усиление персонализации. На текущем этапе система формирует рекомендации, опираясь на оценки пользователей по предметам,

оценки образовательных мероприятий, а также их взаимодействие с мероприятиями (записи и завершения). В будущем можно внедрить более сложные подходы к персонализации, включая:

- Глубокий анализ исторических данных. Расширение возможностей модели за счет более детального учета истории взаимодействий пользователей с системой. Например, анализ частоты и последовательности записей и завершений образовательных мероприятий позволит создавать более точные профили интересов и предлагать наиболее подходящие рекомендации.
- Использование контекстуальных данных. Учет внешних факторов, таких как географическое положение, время года, выбранный формат обучения (онлайн или офлайн), может значительно повысить качество рекомендаций, позволяя адаптировать их под текущие условия и потребности пользователей в режиме реального времени.

Адаптация к динамическим изменениям предпочтений пользователей

Одной из ключевых задач рекомендательных систем является способность быстро адаптироваться к изменениям интересов и предпочтений пользователей. Для решения этой проблемы можно использовать следующие подходы:

- Применение алгоритмов машинного обучения для динамической настройки. Использование технологий машинного обучения, таких как обучение с подкреплением или онлайн-обучение, позволит системе обновлять свои рекомендации на основе поступающих данных, обеспечивая их актуальность в режиме реального времени.
- Сбор пользовательской обратной связи. Внедрение инструментов для получения отзывов от пользователей - через оценки, комментарии или опросы - поможет не только повысить точность рекомендаций, но и своевременно адаптировать модель под изменяющиеся предпочтения и запросы.

Углубление анализа сходства между мероприятиями

Для повышения точности рекомендаций в области образовательных мероприятий важно развивать подходы к анализу их сходства. В перспективе можно реализовать следующие направления:

- Разработка многозадачных и многоуровневых моделей. Применение современных методов глубокого обучения, включая нейронные и сверточные сети, для многогранной оценки схожести мероприятий. Это позволит учитывать не только их текстовые описания, но и такие аспекты, как образовательные цели, контекст проведения и предпочтения целевой аудитории.
- Расширение анализа контекстных характеристик мероприятий. Учет дополнительных данных, таких как отзывы о преподавателях, сложность курсов, продолжительность и формат обучения. Такой подход обеспечит более точные рекомендации, адаптированные к ожиданиям и потребностям пользователей.

Использование новых подходов и технологий

Для дальнейшего совершенствования гибридной рекомендательной системы в будущем можно использовать современные технологии и подходы, такие как:

- Применение генеративных моделей. Использование таких технологий, как Generative Adversarial Networks (GANs), для создания рекомендаций, способных не только учитывать сходство с текущими мероприятиями, но и предлагать неожиданные, инновационные варианты, которые пользователи могли бы упустить, используя стандартные механизмы фильтрации.
- Повышение прозрачности и интерпретируемости. Разработка инструментов, объясняющих пользователям причины появления тех или иных рекомендаций. Это позволит укрепить доверие к системе и обеспечит более глубокое понимание принципов её работы, делая взаимодействие с ней более комфортным и эффективным.

Расширение гибридных подходов

Текущая система применяет гибридный подход, объединяя различные методы для создания рекомендаций. В будущем можно добавить следующие подходы для дальнейшей оптимизации:

- Гибридизация с использованием мета-данных. Включение мета-данных, таких как категория курса, уровень вовлеченности преподавателя, место проведения, что позволит создавать более точные модели, учитывающие разнообразные параметры образовательных мероприятий.
- Гибридизация с применением сетевого анализа. Использование методов сетевого анализа для учета взаимосвязей между участниками (например, пользователями, посетившими одни и те же мероприятия), что поможет строить более сложные модели рекомендаций, учитывающие влияние социальных факторов.

Мониторинг и улучшение эффективности

Для постоянного совершенствования системы необходимо внедрить механизм её мониторинга в реальном времени. Это будет включать:

- Оценку качества рекомендаций. Разработка системы для регулярной проверки точности и актуальности предложений, с применением таких метрик, как: Click-through rate (CTR), conversion rate, user satisfaction и других. Это поможет быстро выявлять проблемы и улучшать работу системы.
- Динамическое тестирование. Применение методов A/B тестирования и других подходов для оценки изменений в рекомендательной системе и анализа их влияния на поведение пользователей.

В этой под-главе обсуждаются возможные пути улучшения и дальнейшего развития гибридной рекомендательной системы, ориентированной на образовательные мероприятия. Эффективное совершенствование таких систем требует постоянного учета изменений в технологиях, условиях и предпочтениях пользователей. В перспективе это открывает возможности для внедрения новых методов и подходов, которые

смогут значительно повысить точность, персонализацию и актуальность рекомендаций.

В данном разделе описывалось будущее развитие предлагаемой гибридной рекомендательной системы.

Вывод по четвёртой главе

На основе метрик, описанных в главе 3 было проведено тестирование эффективности, предлагаемой гибридной рекомендательной системы с последующим анализом полученных результатов. Также была описана потенциальная возможность интеграции, предлагаемой гибридной рекомендательной системы в веб-приложение и даны рекомендации по будущим возможным улучшениям системы.

Заключение

В процессе выполнения диссертация была определена актуальность исследования, а также проанализировано назначение рекомендательных систем и подходы к их построению.

Основной частью данной работы является разработка алгоритмов анализа успеваемости школьников на основе данных об оценках для персонализированной выдачи рекомендаций по обучающим мероприятиям.

Для успешной разработки алгоритмов были определены следующие ключевые этапы:

- провести анализ существующих подходов к построению рекомендательных систем;
- провести анализ методов и алгоритмов необходимых для построения рекомендательной системы и на их основе выбрать подходящий подход к построению собственной рекомендательной системы;
- описать механизм работы рекомендательной системы, выдающей персонализированные рекомендации образовательных мероприятий на основании данных об оценках;
- практически реализовать рекомендательную систему основываясь на данных описанного механизма работы;
- протестировать качество выдаваемых рекомендаций и возможность интеграции разработанной рекомендательной системы в сторонние приложения.

В рамках данной работы была разработана гибридная рекомендательная система, которая анализирует успеваемость школьников на основе данных об оценках и предоставляет персонализированные рекомендации по обучающим мероприятиям. Эта система сочетает в себе элементы не персонализированной, контентной и совместной фильтрации, что позволяет формировать более точные и индивидуальные рекомендации, учитывающие уникальные предпочтения каждого пользователя. В результате, система

способствует оптимизации образовательного процесса, создавая условия для повышения общей успеваемости и достижения учащимися более высоких образовательных результатов.

В главе 4 были представлены результаты тестирования гибридной рекомендательной системы с использованием метрик precision@k , recall@k , F-score@k , HR (Hit Rate) и MRR (Mean Reciprocal Rank). Эти метрики позволяют оценить точность, полноту и качество рекомендаций.

Precision@20: Среднее значение precision@20 составило 0.82, что указывает на высокую точность рекомендаций. Это означает, что большинство рекомендованных образовательных мероприятий являются релевантными для пользователей.

Recall@20: Среднее значение recall@20 составило 0.77, что свидетельствует о хорошей полноте рекомендаций. Система успешно охватывает большинство подходящих образовательных мероприятий.

F-score@20: Среднее значение F-score@20 составило 0.76, что является гармоническим средним между precision и recall , подтверждая баланс между точностью и полнотой рекомендаций.

HR (Hit Rate): Значение HR составило 1.00, что указывает на то, что все пользователи получили хотя бы одну релевантную рекомендацию. Это подтверждает полноту системы и её способность находить подходящие образовательные мероприятия для каждого пользователя.

MRR (Mean Reciprocal Rank): Значение MRR составило 0.86, что свидетельствует о том, что в среднем первый релевантный элемент находится на высокой позиции в списке рекомендаций. Это означает, что система эффективно выстраивает порядок рекомендаций, предоставляя пользователям наиболее подходящие мероприятия в начале списка.

В ходе пользовательского тестирования в специально созданной тестовой среде в виде веб-сайта были выявлены следующие ключевые моменты:

- Система успешно учитывает успеваемость ученика при формировании рекомендаций, что позволяет адаптировать предложенные образовательные мероприятия под уровень знаний и подготовки каждого пользователя. Это помогает избежать перегрузки ученика материалом, который он может не в состоянии усвоить на данном этапе обучения.
- В контексте коллаборативной фильтрации была обнаружена проблема разреженности данных, что приводит к трудностям в нахождении похожих пользователей. Эту проблему решает модуль контентной фильтрации, который обеспечивает точность рекомендаций даже при недостатке данных.
- Система предлагает пользователю курсы исключительно в рамках узкой области его интересов, что ограничивает его выбор. Это может привести к тому, что пользователи не получают возможность узнать о других курсах, которые могли бы им понравиться.

Была успешно проведена интеграция гибридной рекомендательной системы в веб-приложение. Система успешно формирует рекомендации на основе данных о пользователях и их взаимодействиях с образовательными мероприятиями. Это подтверждает возможность практического применения разработанных алгоритмов.

При тестировании возможности интеграции рекомендательной системы в веб-приложения и при её функциональном тестировании программных сбоев обнаружено не было.

Разработанную гибридную рекомендательную систему можно использовать на образовательных платформах, таких как: АСУ РСО.

Список используемой литературы и используемых источников

1. Анатомия рекомендательных систем. Часть первая [Электронный ресурс] // Хабр - Режим доступа: <https://habr.com/ru/companies/lanit/articles/420499/> (Дата обращения: 15.01.2024)
2. Гаврилова Ю. Что такое рекомендательные системы и как они работают [Электронный ресурс] // Skillbox - Режим доступа: https://skillbox.ru/media/code/chto_takoe_rekomendatelnye_sistemy_i_kak_oni_rabotayut/?ysclid=lr50zk8jml990523023 (Дата обращения: 10.01.2024)
3. Дмитриев Н.А., Авласевич Д.В., Шаврина О.В., Чураев В.В. БАЗА ДАННЫХ // Форум молодых ученых. 2020. №10 (50). URL: <https://cyberleninka.ru/article/n/baza-dannyh> (дата обращения: 15.01.2024)
4. Захарова И., Осипов А. Рекомендательные системы: что это и как работает алгоритм рекомендаций [Электронный ресурс] // Mindbox - Режим доступа: <https://mindbox.ru/journal/education/rekomendatelnye-sistemy/?ysclid=lr50zm4qu0308335948> (Дата обращения: 10.01.2024)
5. Зачем мне пылесос с ананасом или как оценить корректность рекомендательной системы [Электронный ресурс] // Хабр – Режим доступа: <https://habr.com/ru/articles/779038/> (Дата обращения: 16.01.2024)
6. Зубарев В. Рекомендательные системы [Электронный ресурс] // Вастрик - Режим доступа: <https://vas3k.blog/blog/355/> (Дата обращения: 10.01.2024)
7. Инновации в рекомендательных списках: новые форматы и перспективы использования [Электронный ресурс] // Научные Статьи.Ру – Режим доступа: https://nauchniestati.ru/spravka/novye-formaty-rekomendatelnyh-spiskov-innovaczii-i-perspektivy-ispolzovaniya/?ysclid=lrg6orli5i198265296#Новые_форматы_рекомендательных_списков (Дата обращения: 16.01.2024)

8. Ищем по-соседски: методы приближённого поиска ближайших соседей для А/В-тестирования гипотез [Электронный ресурс] // Хабр - Режим доступа: <https://habr.com/ru/companies/sberbank/articles/726532/> (Дата обращения: 10.06.2024)
9. Как технологии помогают устранять пробелы в знаниях школьников [Электронный ресурс] // Яндекс-Дзен – Режим доступа: <https://dzen.ru/a/Y4dgioU5YjKy0dw-> (Дата обращения: 16.01.2024)
10. Как устроены рекомендательные системы [Электронный ресурс] // Код – журнал Яндекс Практикума - Режим доступа: <https://thecode.media/recommender/?ysclid=lr7sfltzk193338857> (Дата обращения: 10.01.2024)
11. Коллаборативная фильтрация простыми словами [Электронный ресурс] // ЛаЛаЛань - Режим доступа: <https://lala.lanbook.com/kollaborativnaya-filtraciya-prostymi-slovami> (Дата обращения: 11.01.2024)
12. Модели машинного обучения: что это и как выбрать подходящую [Электронный ресурс] // Блог Яндекс практикума – Режим доступа: <https://practicum.yandex.ru/blog/modeli-mashinnogo-obucheniya/> (Дата обращения: 10.06.2024)
13. Персонализация образования: что это такое [Электронный ресурс] // Skillbox – Режим доступа: <https://skillbox.ru/media/education/chto-takoe-personalizatsiya-obrazovaniya/> (Дата обращения: 16.01.2024)
14. Рекомендательная система с нуля [Электронный ресурс] // Skine.ru – Режим доступа: <https://skine.ru/articles/14664/?ysclid=lrav5hkc3z32777356> (Дата обращения: 16.01.2024)
15. Рекомендательные системы [Электронный ресурс] // Сравни - Режим доступа: <https://www.sravni.ru/kursy/info/rekomendatelnye-sistemy/> (Дата обращения: 09.01.2024)
16. Самарина А. Что такое рекомендательные системы и как работают алгоритмы рекомендаций [Электронный ресурс] // Комьюнити - Режим

доступа: <https://timeweb.com/ru/community/articles/chto-takoe-rekomendatelnye-sistemy-i-kak-rabotayut-algoritmy-rekomendaciy> (Дата обращения: 15.01.2024)

17. Системы рекомендаций: обзор современных подходов [Электронный ресурс] // Cyberleninka - Режим доступа: <https://cyberleninka.ru/article/n/sistemy-rekomendatsiy-obzor-sovremennyh-podhodov?ysclid=lx4v6qsrr440286052> (Дата обращения: 10.06.2024)

18. 7 Types of Hybrid Recommendation System [Электронный ресурс] // Medium – Режим доступа: <https://medium.com/analytics-vidhya/7-types-of-hybrid-recommendation-system-3e4f78266ad8> (Дата обращения: 08.06.2024)

19. A collaborative filtering recommendation framework utilizing social networks [Электронный ресурс] // ScienceDirect – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S2666827023000488> (Дата обращения: 10.06.2024)

20. A Hybrid Semantic Recommender System Based on an Improved Clustering [Электронный ресурс] // ResearchGate – Режим доступа: https://www.researchgate.net/publication/374990558_A_Hybrid_Semantic_Recommender_System_Based_on_an_Improved_Clustering (Дата обращения: 18.01.2025)

21. Beginners Guide to learn about Content Based Recommender Engines Python [Электронный ресурс] // Analytics Vidhya – Режим доступа: <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/> (Дата обращения: 10.06.2024)

22. Falk K. Practical Recommender Systems, Manning Publications: 1st edition. – 2020. – 448 p.

23. Introduction to Recommender Systems [Электронный ресурс] // Tryolabs – Режим доступа: <https://tryolabs.com/blog/introduction-to-recommender-systems> (Дата обращения: 10.01.2024)

24. Kompan M., Gaspar P., Bielikova M. Hybrid Collaborative Recommendations: Practical Considerations and Tools to Develop a Recommender [Электронный ресурс] // ResearchGate – Режим доступа:

https://www.researchgate.net/publication/329396549_Hybrid_Collaborative_Recommendations_Practical_Considerations_and_Tools_to_Develop_a_Recommender
(Дата обращения: 08.06.2024)

25. Non-Personalised Recommender System in Python [Электронный ресурс] // Medium – Режим доступа: <https://medium.com/@tomar.ankur287/non-personalised-recommender-system-in-python-42921cd6f971> (Дата обращения: 09.06.2024)

26. Performance Evaluation of Time-based Recommendation System in Collaborative Filtering Technique [Электронный ресурс] // ScienceDirect – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S1877050923001618> (Дата обращения: 10.06.2024)

27. Ricci, F., Rokach, L., & Shapira, B. Recommender Systems Handbook / F. Ricci, L. Rokach, B. Shapira // Springer. – 2022. – 3rd ed.

28. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison [Электронный ресурс] // ScienceDirect – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S1319157821002652> (Дата обращения: 10.06.2024)

29. Unit IV Knowledge and Hybrid Recommendation System [Электронный ресурс] // Slideshare – Режим доступа: <https://www.slideshare.net/slideshow/unit-iv-knowledge-and-hybrid-recommendation-systempdf/267037336> (Дата обращения: 08.06.2024)

30. What is Gradient Descent? the “Learning” in Machine Learning [Электронный ресурс] // Medium – Режим доступа: <https://medium.com/@saidisouhaiebai/what-is-gradient-descent-the-learning-in-machine-learning-63881f8845b4> (Дата обращения: 10.06.2024)